CLINICAL DECISION SUPPORT FOR THE WHOLE GENOME SEQUENCE

by

Brandon Marshall Welch

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biomedical Informatics

The University of Utah

May 2014

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The following faculty members served as the supervisory committee chair and members for the dissertation of Brandon Marshall Welch. Dates at right indicate the members' approval of the dissertation.

Kensaku Kawamoto, Chair	4/16/2014 Date Approved
Karen Eilbeck, Member	4/18/2014 Date Approved
Guilherme Del Fiol, Member	4/16/2014 Date Approved
Joshua D. Schiffman, Member	4/18/2014 Date Approved
Michael W. Varner, Member	 Date Approved

The dissertation has also been approved by Wendy W. Chapman Chair of the Department of Biomedical Informatics and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

The widespread use of genomic information to improve clinical care has long been a goal of clinicians, researchers, and policy-makers. With the completion of the Human Genome Project over a decade ago, the feasibility of attaining this goal on a widespread basis is becoming a greater reality. In fact, new genome sequencing technologies are bringing the cost of obtaining a patient's genomic information within reach of the general population. While this is an exciting prospect to health care, many barriers still remain to effectively use genomic information in a clinically meaningful way. These barriers, if not overcome, will limit the ability of genomic information to provide a significant impact on health care. Nevertheless, clinical decision support (CDS), which entails the provision of patient-specific knowledge to clinicians at appropriate times to enhance health care, offers a feasible solution. As such, this body of work represents an effort to develop a functional CDS solution capable of leveraging whole genome sequence information on a widespread basis. Many considerations were made in the design of the CDS solution in order to overcome the complexities of genomic information while aligning with common health information technology approaches and standards. This work represents an important advancement in the capabilities of integrating actionable genomic information within the clinical workflow using health informatics approaches.

I dedicate this to my loving wife, Marie, who lived in ‘the desert’ for over three years and endured two pregnancies while I completed my PhD work. Also to my parents who have always been supportive of my educational pursuits.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge the members of my PhD committee: Dr. Kawamoto, Dr. Eilbeck, Dr. Del Fiol, Dr. Schiffman, and Dr. Varner for their valuable contribution and guidance throughout my PhD efforts. I'd particularly like to thank Dr. Kawomoto for giving me the many opportunities and allowing me to excel during my time as a PhD student. I'd also like to thank Drs. Schiffman, Varner, Logan, and Brixner from the Program in Personalized Health Care for bringing me into their program as a Predoctoral Fellow which funded my PhD efforts and gave me many more opportunities to leverage my skills and talents. Also to the many students, faculty, and staff who have contributed ideas to my dissertation. Finally, thanks to the journals that have reviewed and published the works that make up this dissertation.

INTRODUCTION

Genetics in medicine began in the early 20th century with physicians observing certain metabolic disorders re-occurring in families that could be explained by Mendel's laws of inheritance. Over the next century, as the scientific understanding of genetics grew, its application in medicine has seen slow but steady growth, mostly in rare, single-gene disorders. However, with the recent rapid advances in genetics research and technology, fueled by the Human Genome Project, our understanding of genetics to human disease is rapidly becoming more ubiquitous across the specialties of medicine. Indeed, Dr. Francis Collins, the current Director of the National Institutes of Health, has stated that "virtually every disease has some genetic component."

Understanding the genetic contribution to disease etiology and individualized drug therapy has given rise to the field of personalized medicine, which is defined as the tailoring of medical treatment to the individual characteristics of each patient. The ability to personalize care to the individual patient is an exciting prospect to many in health care, as it has the potential to reduce costs and improve the quality of care. Personalized care is the provision of health based upon the individual characteristics, including genotype, of the patient. Characteristics used to provide personalized health care can include personal health history, genotype, biomarkers, environmental and social influences, and other attributes that can be used to classify or subtype an individual for personalized care. For this

dissertation, the primary focus is on the application of an individual's genomic information to provide personalized health care.

Given the potential of genomic information to enable personalized health care, its ability to be effectively leveraged in a clinical setting is important. However, many challenges exist which inhibit such effective clinical application of genomic information. These challenges will be highlighted throughout subsequent chapters. Nevertheless, we believe that clinical decision support (CDS) is an effective mechanism to overcome these challenges and support the provision of genomic information to support personalized medicine. CDS entails providing clinicians, patients, and other healthcare stakeholders with pertinent knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and healthcare.

Our first step in this research was to identify the body of work related to CDS for genetically-guided personalized medicine. For this particular effort we conducted a systematic review of the published literature on the topic to identify and learn from trends and patters. The work resulted in a *Journal of the American Medical Informatics Association* (JAMIA) publication entitled "Clinical decision support for genetically guided personalized medicine: a systematic review" and is found in Chapter 1 of this dissertation. A key finding from this research is that we identified zero publications on CDS capabilities for whole genome sequence (WGS) information. As a result, this set us on the course of researching and developing CDS capabilities for WGS information.

To build the case of the need for CDS capabilities for WGS information, we authored a publication entitled "The need for clinical decision support integrated with the electronic health record for the clinical application of whole genome sequencing

information" which was published in the *Journal of Personalized Medicine*. This work, found in Chapter 2 of this dissertation, describes in detail the barriers of effectively using WGS information in a clinical setting. This work further describes how CDS can support WGS information in an effective way and illustrates several clinical scenarios in which this type of CDS could be leveraged.

Subsequently, during the design process of a CDS architecture for WGS information, we identified several important requirements for such an architecture to be effective which were not identified in previous research. As a result, we developed these requirements into a formal desiderata and validated them among domain experts in genomics and CDS. This work entitled "Technical desiderata for the integration of genomic data with clinical decision support" is currently under review for publication. This work is represented in Chapter 3 of this dissertation.

Following the development and validation of this desiderata, we designed an architecture approach that could satisfy the requirements of the desiderata. Indeed, this approach leveraged service-oriented architecture (SOA) design principles separate out business concerns to independently manage and control components or services. This approach allowed the architecture the scalability and flexibility to adapt to the challenges in providing WGS information to support the provision of personalized health care. This architecture design along with design justifications are describe in the manuscript entitled "A proposed clinical decision support architecture capable of supporting whole genome sequence information" in the *Journal of Personalized Medicine* and makes up Chapter 4 of this dissertation.

Finally, to prove that this architectural approach is a feasible solution, we developed

and tested a prototype patterned after the proposed architecture described above. In this effort, we were able to successfully develop this prototype using open source solutions and health information technology standards. While some of these components used required significant configuration or development, this effort demonstrates it is possible to use this approach to provide WGS-guided CDS at the point of care within the clinical workflow. This work is currently under review for publication and is described in Chapter 5.

In summary, this work represents an advance in informatics solutions and approaches that can support genome-guided personalized medicine through CDS capabilities. While it may be several years before this approach can be used on a widespread basis and impact the general population, it provides a foundation upon which further research and development can be based. In fact, many opportunities exist for ongoing work supported by grant funding opportunities and for further journal publications in this area. Nevertheless, the ultimate goal is for these CDS solutions to be implemented in the clinical setting so that they can improve patient outcomes through personalized medicine.

CHAPTER 1

CLINICAL DECISION SUPPORT FOR GENETICALLY GUIDED PERSONALIZED MEDICINE: A SYSTEMATIC REVIEW

Reprinted with permission from:

Welch BM, Kawamoto K. Clinical decision support for genetically guided personalized medicine: a systematic review. *J Am Med Inform Assoc*. 2013 Mar-Apr;20(2):388-400.

Review

Clinical decision support for genetically guided personalized medicine: a systematic review

Brandon M Welch,[1] Kensaku Kawamoto[2]

ABSTRACT

Objective To review the literature on clinical decision support (CDS) for genetically guided personalized medicine (GPM).

Materials and Methods MEDLINE and Embase were searched from 1990 to 2011. The manuscripts included were summarized, and notable themes and trends were identified.

Results Following a screening of 3416 articles, 38 primary research articles were identified. Focal areas of research included family history-driven CDS, cancer management, and pharmacogenomics. Nine randomized controlled trials of CDS interventions for GPM were identified, seven of which reported positive results. The majority of manuscripts were published on or after 2007, with increased recent focus on genotype-driven CDS and the integration of CDS within primary clinical information systems.

Discussion Substantial research has been conducted to date on the use of CDS to enable GPM. In a previous analysis of CDS intervention trials, the automatic provision of CDS as a part of routine clinical workflow had been identified as being critical for CDS effectiveness. There was some indication that CDS for GPM could potentially be effective without the CDS being provided automatically, but we did not find conclusive evidence to support this hypothesis.

Conclusion To maximize the clinical benefits arising from ongoing discoveries in genetics and genomics, additional research and development is recommended for identifying how best to leverage CDS to bridge the gap between the promise and realization of GPM.

► An additional appendix is published online only. To view this file please visit the journal online (http://dx.doi.org/10.1136/amiajnl-2012-000892).

[1]Department of Biomedical Informatics and Program in Personalized Health Care, University of Utah, Salt Lake City, Utah, USA
[2]Department of Biomedical Informatics and Information Technology Services, University of Utah, Salt Lake City, Utah, USA

Correspondence to
Dr Kensaku Kawamoto, Department of Biomedical Informatics, University of Utah, 615 Arapeen Way, Suite 208, Salt Lake City, UT 84092, USA; kensaku.kawamoto@utah.edu

Received 10 February 2012
Accepted 27 July 2012
Published Online First 22 August 2012

BACKGROUND

Genetically guided personalized medicine (GPM) entails the delivery of individually tailored medical care that leverages information about each person's unique genetic characteristics.[1] The promise of GPM has expanded as advances in genomics have accelerated over the past several decades. This promise of GPM is that research discoveries will one day lead to medical treatments and therapies that are tailored to the individual characteristics of each patient, including clinical data, genetic test results, patient preference, and family health history (FHx). GPM has the potential to increase the efficacy, quality, and value of healthcare by providing individually optimized prevention, diagnosis, and treatment.[2]

As ongoing research continues to expand the GPM knowledge base, it has become increasingly important to translate this knowledge into routine healthcare practice in order to realize the promise of GPM.[3] However, the effective realization of GPM remains very limited.[4] While this is partly due to the need for further evidence of the clinical utility and cost effectiveness of a genetically guided approach to patient care, an important additional reason is the need for information systems that assist in the translation of knowledge from bench to bedside.[5] Even without the complexity of genetics, it can often take over 15 years to translate research from bench to bedside.[6] This translational bottleneck is likely to be an even more significant problem in GPM for the following reasons.

Limited genetic proficiency of clinicians

Many clinicians receive minimal training in clinical genetics. As a result, many physicians lack the confidence and understanding needed for effectively interpreting and using genetic information in their clinical practices.[7]

Limited availability of genetics experts

Currently, there are about 3000 board-certified genetic counselors[8] and approximately 1200 medical geneticists practicing in the USA (S. R. DelBusso, American Board of Medical Genetics Administrator, October 28, 2011, personal communication). The growing utility of genetic information is putting an increasing burden on these professionals. We cannot expect these genetics experts to be readily available each time genetic information should be used to guide medical treatment. For effective, efficient, and widespread clinical use, the burden of genetic interpretation and guidance must be shared by the wider clinical community.

Breadth and growth of genetic knowledge base

There are currently over 2500 clinical genetic tests available to clinicians, encompassing a wide breadth of medical care.[9] It is therefore unreasonable to expect a clinician to remember every appropriate genetic test for a particular condition in conjunction with test-specific guidelines for ordering and interpretation. Compounding this issue, the continual growth in the knowledge base and the prospect of full genome sequencing will inevitably overwhelm clinicians' capacities to manage and leverage this information effectively for GPM unless computerized assistance is provided for interpreting and acting on this information.

Various investigators and leaders have identified health information technology as being vital to overcoming these barriers and realizing the promise of GPM.[2 10] In particular, clinical decision support (CDS) has been identified as a critical enabler of GPM.[11 12] CDS entails providing clinicians, patients, and other healthcare stakeholders with pertinent knowledge and person-specific

information, intelligently filtered or presented at appropriate times, to enhance health and healthcare.[13] CDS has the capacity to process complex, disparate data and present actionable, standardized, evidence-based recommendations in a way that is usable by a clinician in everyday practice.[11] As such, CDS can help bridge the gap between the promise and realization of GPM (figure 1). Given the criticality of CDS for realizing the promise of GPM, and given the lack of a systematic review on this topic, we sought in this paper to assess the history and state of CDS for GPM through a systematic review of the literature.

METHODS

Data sources and inclusion criteria

We searched MEDLINE and Embase from 1990 to 2011 using a search strategy adapted from previous systematic reviews of CDS,[14] genetic health services,[15] and FHx[16] (see supplementary appendix, available online only, for full search strategy). The final literature search was conducted on June 1, 2012. The inclusion criteria for the review were as follows: English article; human focus; manuscript in peer-reviewed journal; and primary focus on the use of computers to deliver genetically guided, patient-specific assessments and/or recommendations to healthcare providers and/or patients to guide clinical decision-making, as further defined in Box 1.

For all identified references, the authors reviewed titles, index terms, and available abstracts to determine if the articles appeared to meet all inclusion criteria. If insufficient information was available to make a confident decision at this stage, the article was included for full-text retrieval. Each full-text article was then reviewed to determine its final inclusion status.

Data abstraction

For each of the articles that met the inclusion criteria listed above, we abstracted data on the clinical application area, CDS type, genetic information used, primary users, article type, study location, CDS purpose, and notable informatics aspects. CDS type was defined as being either stand-alone CDS or integrated CDS. A stand-alone CDS system is a CDS system that exists in isolation from a primary clinical information system containing relevant patient data, such as an electronic health record (EHR) system. A stand-alone CDS system requires manual data input before a CDS result can be produced. In contrast, an integrated CDS system is integrated with a primary clinical information system such as an EHR system or a computerized provider order entry system to aggregate necessary patient-specific information automatically and to provide guidance within routine clinical workflows. Clinical application area was defined as the clinical domain targeted by the CDS intervention. Article type consisted of system description papers and evaluation studies of various types (eg, qualitative evaluation, randomized controlled trial). Genetic information used consisted of FHx, genotype, or both. Primary users were defined as the individuals who primarily entered information and received the results. Study location was the country or region where the research was conducted. CDS purpose identified the role of the CDS system within the context of clinical decision-making. A notable informatics aspect was also abstracted if a manuscript utilized a methodology that was considered to be of potential interest to an informatics audience. For intervention studies, additional details regarding the study size and study outcomes were abstracted.

Data analysis and presentation

Using the abstracted attributes, the manuscripts were grouped into logical categories, primarily according to CDS type and clinical application area. The findings from these manuscripts were summarized through tables and narrative discussion. In addition, notable themes and trends were identified and discussed. A quantitative analysis of CDS trials to identify features predictive of trial outcomes was considered.[14] However, due to the limited sample size of CDS trials available, such a quantitative analysis of potential success factors was not feasible.

RESULTS

The initial MEDLINE and Embase searches identified 3416 potentially relevant articles. During the title and abstract review, 82 articles were rejected for not being in English, 504 articles were rejected because they were not focused on humans, 34 articles were rejected for not being a peer-reviewed manuscript, and 2494 articles were rejected because the primary focus of the work was not on the use of computers to deliver genetically guided, patient-specific assessments and/or recommendations. The remaining 302 articles underwent full-text review, at which stage 37 articles were rejected for not being a peer-reviewed primary research article and 227 articles were rejected because the primary focus of the work was not on the use of computers

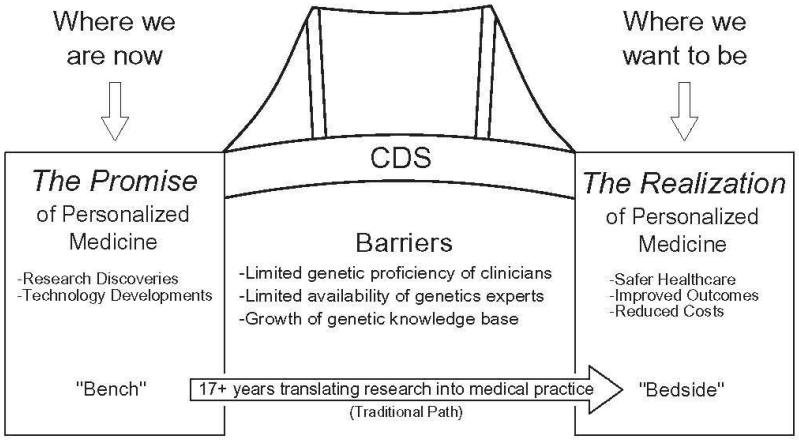


Figure 1 Clinical decision support (CDS) as bridge overcoming barriers to genetically guided personalized medicine.

Box 1 Manuscript inclusion criteria

- Definitions:
  - Healthcare provider = physician, nurse practitioner, physician assistant, registered nurse, or genetic counselor
  - Genetic factor = genotype, gene expression profile, and/or family health history
- Universal inclusion criteria:
  - English article
  - Human focus
  - Manuscript in peer-reviewed journal
- Additional inclusion criteria (at least one):
  - Intervention study evaluating the impact of a CDS system in an actual patient care context
    - For a comparative intervention study, CDS required to be a part of the primary intervention under evaluation
    - Excludes laboratory evaluations or simulation studies
  - Methodology article whose primary focus is on how CDS systems should be designed specifically to support clinical delivery of patient-specific assessments and/or recommendations guided by genetic factors. Includes system description articles.

to deliver genetically guided, patient-specific care guidance (figure 2). The final set of included manuscripts consisted of 38 primary research articles.[17–54] The manuscripts included were published from 1990 to 2011, with the majority of manuscripts published on or after 2007. Provided below is a summary and analysis of these earlier works, grouped primarily by CDS type and area of clinical focus.

CDS systems for genetically guided cancer management

Genetically guided cancer management was the focus of 22 primary research articles summarized in tables 1–4.[17–37 54] These manuscripts include six manuscripts related to the Risk Assessment in Genetics (RAGs) system for providing FHx-driven CDS (table 1),[17–21 54] six manuscripts on other FHx-driven CDS tools for breast cancer management (table 2),[22–27] four manuscripts on genotype-driven CDS tools for breast cancer management (table 3),[28 31] and six additional manuscripts on GPM CDS tools for non-breast cancer management (table 4).[32 37]

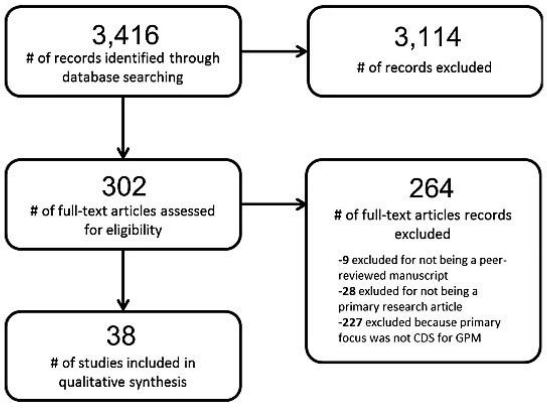


Figure 2 Manuscript selection process. CDS, clinical decision support; GPM, genetically guided personalized medicine.

RAGs system for providing FHx-driven CDS

Some of the earliest and most comprehensive research on the use of CDS to support GPM was conducted by Emery[55] (table 1), who identified that existing systems were not designed for primary care and that none provided patient management advice based on calculated risk. To address this gap, Emery developed a system known as RAGs, which helped general practitioners (GPs) in the UK collect FHx relevant to familial breast, ovarian, and colorectal cancer and provided appropriate management guidance, primarily regarding guideline-based specialist referrals.[17–19 54] A later extension of the RAGs system was referred to as the GRAIDS system.[20 21] This body of work included several favorable evaluations of these systems,[18 19 21] including a cluster randomized controlled trial (RCT) across 45 GP teams that found that GRAIDS significantly increased the proportion of patients referred appropriately to the regional genetics clinic according to evidence-based practice guidelines.[21]

Other FHx CDS tools for breast cancer management

Beyond the work of Emery,[55] CDS research for GPM has focused heavily on breast cancer management (table 2). Risk assessment tools for breast cancer can enable personalized care according to an individual's level of risk.[22 25] An RCT conducted in the UK found that a stand-alone breast cancer CDS tool had limited impact due to lack of awareness and use by GPs.[24] At the same time, a stand-alone CDS tool that calculated risks for breast cancer, heart disease, osteoporosis, and endometrial cancer was shown in an RCT to enhance the effectiveness of genetic counselors using the system.[25 26] Another stand-alone CDS system that has been found to be beneficial is HughesRiskApps, which collects relevant FHx information and provides clinicians with various tools to support the management of patients. An observational implementation study of this tool in a community hospital setting found significant adoption and impact.[27]

Genotype-driven CDS tools for breast cancer management

Several investigators have developed CDS systems that support treatment and decision-making once mutations have been identified in the breast cancer (BRCA) genes (table 3). In the UK, Glasspool and colleagues[30 31] developed a CDS tool known as REACT (Risks, Events, Actions and their Consequences over Time), which used a graphical timeline display to model real-time changes in lifetime risks as a result of risk-reduction interventions for breast cancer and ovarian cancer. In addition, several patient-directed, stand-alone CDS systems have been developed for improving risk communication and decision-making in breast cancer management based on BRCA genotype.[28 29]

CDS for other cancers

Besides breast cancer, other cancers have been the focus of CDS research and development (table 4). Most of this CDS research for other cancers has involved colorectal cancer, and in particular Lynch syndrome—a strongly heritable type of colorectal cancer.[32 34] Of note, the RAGs and GRAIDS systems described earlier supported both breast cancer and colorectal cancer management.[17–21 54] An additional CDS system investigated for colorectal cancer management is CRCAPRO, similar to BRCAPRO, which used FHx to identify patients at risk of hereditary colorectal cancer.[33] In addition, a group in the Netherlands developed a CDS intervention to remind pathologists to order Lynch syndrome genetic testing among patients

Table 1 Summary of primary research on CDS systems for cancer-related GPM: RAGs system for providing FHx-driven CDS

Citation and name of system (if applicable)	Manuscript summary and trial details (if applicable)	Users and study location	Genetic information used	Integrated with primary clinical information system	CDS purpose and clinical focus	Manuscript type	Notable informatics aspect
Coulson, 2001[17]; RAGs	System description of RAGs, which was designed to help GPs build a family pedigree, calculate genetic risk, and obtain guideline-based care recommendations	GPs in UK	FHx	No	Assessment of patient risk and provision of management recommendations for familial breast, ovarian, and colorectal cancer	System description	RAGs uses PROforma, an argumentation-based technology for CDS
Emery, 1999[18]; RAGs	A qualitative evaluation of RAGs, which found that the system was easy to use by GPs and served as an appropriate application of information technology to assist with clinical care	GPs in UK	FHx	No	Assessment of patient risk and provision of management recommendations for familial breast, ovarian, and colorectal cancer	Qualitative study	RAGs uses PROforma, an argumentation-based technology for CDS
Glasspool, 2001[54]; RAGs	System description of RAGs, which uses an argumentation approach to assess genetic risk and provide detailed qualitative explanation with referral advice	GPs in UK	FHx	No	Assessment of patient risk and provision of management recommendations for familial breast, ovarian, and colorectal cancer	System description	RAGs uses PROforma, an argumentation-based technology for CDS
Emery, 2000[19]; RAGs	Comparative analysis of RAGs, Cyrillic (a commercially available pedigree drawing system), and pen and paper for use by GPs. This study demonstrated that RAGs significantly improved pedigree accuracy and produced more appropriate management decisions than the other two methods. Furthermore, 92% of GPs preferred RAGs to the other methods	GPs in UK	FHx	No	Assessment of patient risk and provision of management recommendations for familial breast, ovarian, and colorectal cancer	Comparative study	RAGs uses PROforma, an argumentation-based technology for CDS
Emery, 2005[20]; GRAIDS	System description of GRAIDS, a next-generation FHx CDS tool that built on both RAGs and Cyrillic and provided an enhanced user interface for GPs to assess familial cancer risk	GPs in UK	FHx	No	Assessment of patient risk and provision of management recommendations for familial breast, ovarian, and colorectal cancer	System description	Server-based application that provides both heuristic and statistical risk assessment
Emery, 2007[21]; GRAIDS	A cluster RCT of GRAIDS conducted across 45 GP teams in the UK. GRAIDS significantly increased the number of referrals to the regional genetics clinic ($p=0.001$), with the referrals being significantly more likely to be consistent with referral guidelines ($p=0.006$). Moreover, patients referred from GRAIDS practices had significantly lower cancer worry scores at the point of referral ($p=0.02$)	GPs in UK	FHx	No	Assessment of patient risk and provision of management recommendations for familial breast, ovarian, and colorectal cancer	RCT	Server-based application that provides both heuristic and statistical risk assessment

CDS, clinical decision support; FHx, family health history; GP, general practitioner; GPM, genetically guided personalized medicine; GRAIDS, Genetic Risk Assessment in an Intranet and Decision Support; RAGs, Risk Assessment in Genetics; RCT, randomized controlled trial.

Table 2 Summary of primary research on CDS systems for cancer-related GPM: other FHx CDS tools for breast cancer management

Citation and name of system (if applicable)	Manuscript summary and trial details (if applicable)	Users and study location	Genetic information used	Integrated with primary clinical information system	CDS purpose and clinical focus	Manuscript type	Notable informatics aspect
Tsouskas, 1997[22]	Evaluation of a CDS tool that used patient-specific breast cancer risk information, including FHx, to identify patients at high risk of breast cancer. The system identified nine out of 10 women with breast cancers in this study	Clinicians in Europe	FHx	No	Assessment of patient risk for breast cancer	Validation study	Expert system developed using a variant of the BASIC programming language
Berry, 2002[23]; BRCAPRO	Evaluation of BRCAPRO, which predicted the probability of carrying a BRCA mutation based on a patient's FHx. BRCAPRO was effective in predicting the probability of carrying the BRCA mutation	Clinicians in USA	FHx	No	Assessment of patient risk for breast cancer	Validation study	Probability calculated using Bayesian updating
Wilson, 2006[24]	RCT of a stand-alone breast cancer CDS tool to guide referrals in everyday GP practices. The study consisted of 86 GP practices. The CDS system did not result in a statistically significant improvement, due largely to the limited awareness and adoption of the tool by GPs	GPs in UK	FHx	No	Assessment of patient risk and provision of management recommendations for breast cancer	RCT	The deployment of the CDS system was purposely pragmatic and did not involve extensive workflow integration measures
Matloff, 2007[26]	System description of a tool to provide patient-specific predictions of women's future risks for breast cancer, heart disease, osteoporosis, and endometrial cancer utilizing personal and FHx	Genetic counselors in USA	FHx	No	Assessment of patient risk for breast cancer	System description	System used a Markov model
Matloff, 2006[25]	RCT of a CDS tool used by genetic counselors[26] to enable personalized risk assessment and genetic counseling. The trial involved 48 cancer-free, post-menopausal women with a first-degree relative of breast cancer who were contemplating the use of alternative menopausal therapy options. This trial found that patients in the intervention group had increased knowledge and a lower, more accurate perceived risk of developing breast cancer compared to the control group	Genetic counselors in USA	FHx	No	Assessment of patient risk for breast cancer	RCT	System used a Markov model
Ozanne, 2009[27]; Hughes Risk Apps	System description of HughesRiskApps and evaluation of its impact at a community hospital. The CDS system significantly increased the number of patients seen for risk consultation and genetic test ordering. The implementation improved efficiency in several ways and did not require significant investment in capital or personnel	Clinicians in USA	FHx	No	Assessment of patient risk and provision of management recommendations for breast cancer	System description; pre—post comparison	Used tablet computers to collect information from patients. Used Health Level 7 compliant information models.

CDS, clinical decision support; FHx, family health history; GP, general practitioner; GPM, genetically guided personalized medicine; RCT, randomized controlled trial.

who met certain criteria, one of which was a suspicious FHx. This intervention significantly improved pathologists' recognition of patients at risk of Lynch syndrome.[34] Moreover, Dr Henry Lynch, for whom Lynch syndrome is named, developed a CDS system for supporting his hereditary cancer consulting service. This CDS system expedited clinicians' decision-making processes and resulted in a significant reduction in time spent on cases.[32]

Similar to the stand-alone CDS systems for breast cancer management described earlier,[28 29] stand-alone CDS tools have been shown to be useful for the management of other types of cancers, including prostate cancer[37] and alcohol-related

Table 3 Summary of primary research on CDS systems for cancer-related GPM: genotype-driven CDS tools for breast cancer management

Citation and name of system (if applicable)	Manuscript summary and trial details (if applicable)	Users and study location	Genetic information used	Integrated with primary clinical information system	CDS purpose and clinical focus	Manuscript type	Notable informatics aspect
Schwartz, 2009[28]	RCT of patient-facing tool that captured patient-specific information and provided tailored content about risks, benefits and management options based on the patients' particular situations. This study found that among 214 BRCA-positive women who were initially undecided about how to manage their breast cancer risk, patients who used the CDS tool were more likely to reach a management decision (p=0.001), had decreased decision conflict (p=0.002), and increased satisfaction (p=0.002) compared to women who did not use the CDS tool	Patients in USA	Genotype	No	Provision of management recommendations for breast cancer	RCT	CD-ROM based, patient-directed decision aid
Hooker, 2011[29]	Longitudinal RCT of patient-facing BRCA decision aid.[28] This study showed significantly higher cancer-specific distress (p=0.01) and genetic testing-specific distress (p=0.01) among users of the personalized decision aid after one month. Distress levels between groups were the same after 12 months	Patients in USA	Genotype	No	Provision of management recommendations for breast cancer	RCT	CD-ROM based, patient-directed decision aid
Glasspool, 2007[30]; REACT	System description of REACT (Risks, Events, Actions and their Consequences over Time), a breast cancer CDS tool with a graphical timeline display to model real-time changes in lifetime risk as a result of risk-reduction interventions such as tamoxifen therapy, hormone therapy, and mastectomy	Genetic counselors in UK	Genotype	No	Prediction of response to treatment for breast and ovarian cancer	System description	Graphical display of risk changes dynamically based on selected interventions
Glasspool, 2010[31]; REACT	Qualitative study of REACT by eight genetic counselors.[30] Most counselors found REACT effective for genetic risk management, although there were concerns related to the tool's potential to alter the dynamics of the clinician–patient interaction	Genetic counselors in UK	Genotype	No	Prediction of response to treatment for breast and ovarian cancer	Qualitative study	Graphical display of risk changes dynamically based on selected interventions

CDS, clinical decision support; GPM, genetically guided personalized medicine; RCT, randomized controlled trial; REACT, Risks, Events, Actions and their Consequences over Time.

cancers.[36] These studies included an RCT that showed that a patient-directed, genotype-driven CDS tool for alcohol-related cancer risk significantly reduced alcohol consumption by patients at increased genetic risk.[36] These studies, as well as the previous studies on breast cancer,[28 29] showed that patient-directed CDS systems can be clinically useful.

CDS for pharmacogenomics

Pharmacogenomics, the practice of tailoring drug therapy to the patient's unique genetic characteristics, can be a complicated process; genetically guided CDS offers a solution for simplifying this process. Table 5 summarizes the six primary research articles identified on this topic.[38 43] These studies include a description and validation of a CDS system for genetically guided treatment of HIV infections,[38] as well as an RCT that found that genotyping combined with CDS-guided therapy improved outcomes over standard of care.[39] Outside of HIV therapy, other investigators focused on how CDS for pharmacogenomics could be integrated with primary clinical information systems such as computerized provider order entry systems.[40 42 43] These studies evaluated considerations such as developing the underlying pharmacogenomics knowledge base,[40] representation of genetic information in the EHR for supporting pharmacogenomics CDS,[42] and the availability of patient data required for pharmacogenomics within the EHR.[43] The lone stand-alone system for pharmacogenomics used genotype and clinical data to estimate and graphically represent a patient's plasma warfarin concentration over time.[41]

Other CDS systems for GPM

Table 6 summarizes the 10 primary research articles that were neither cancer specific nor focused on pharmacogenomics.[44 53] As with CDS for cancer, there has been a substantial focus on FHx-driven CDS for other medical conditions. For example, a tool called GenInfer considered FHx and calculated inheritance risks for genetic diseases,[44] and FHx-driven CDS was included as a part of the National Russian Genetic Register.[45 46] Beyond these system descriptions, recent studies of FHx-driven CDS have focused on impact evaluation, with mixed results.[48 49 51]

Finally, there were four primary research studies on genotype-driven CDS systems not focused on pharmacogenomics or

Table 4 Summary of primary research on CDS systems for cancer-related GPM: CDS for other cancers

Citation and name of system (if applicable)	Manuscript summary and trial details (if applicable)	Users and study location	Genetic information used	Integrated with primary clinical information system	CDS purpose and clinical focus	Manuscript type	Notable informatics aspect
Evans, 1995[32]	Description of a FHx CDS system developed for a hereditary cancer consulting service. The system collected FHx information, evaluated the FHx for familial risk patterns, and produced preliminary risk assessment and management recommendations. The system resulted in a significant reduction in time spent on cases	Hereditary cancer consulting service in USA	FHx	No	Assessment of patient risk and provision of management recommendations for hereditary cancer	System description; impact observation	Expert rule-based system that modeled the pattern recognition capabilities of clinical geneticists
Bianchi, 2007[33]; CRCAPRO	Evaluation of CRCAPRO, which used FHx of colorectal and endometrial cancers to identify patients with Lynch syndrome. This study showed that CRCAPRO has low sensitivity and specificity	Clinicians in UK	FHx	No	Assessment of patient risk for colorectal cancer	System validation	Probability calculated using Bayesian updating
Overbeek, 2010[34]	RCT of electronic reminders to pathologists to consider Lynch syndrome genetic testing among newly diagnosed colon cancer patients based on FHx. The CDS reminder intervention in 12 pathology laboratories significantly improved pathologists' recognition of patients at risk for Lynch syndrome (OR 2.8; 95% CI 1.1 to 7.0) and increased use of genetic testing (OR 4.1; 95% CI 1.3 to 13.2)	Pathologists in Europe	FHx	Yes	Provision of management recommendations for colorectal cancer	RCT	Electronic reminders provided through health information system
Picone, 2011[35]; NeoMark	System description of NeoMark, a web-based tool that combined medical images, genetic markers, and other patient data before and after treatment of oral cavity squamous cell carcinoma to predict reoccurrence	Clinicians in Europe	Genotype	No	Assessment of patient risk for oral cancer	System description	Uses a service-oriented, modular architecture
Hendershot, 2010[36]	RCT of a web-based genetic feedback intervention involving 200 college students of Asian descent. The system provided personalized alcohol-related health risk information and feedback based on the patient's genotype. The tool resulted in significant reductions in drinking ($p=0.02$) among participants with the genotype associated with higher risk of alcohol-related cancer	Patients in USA	Genotype	No	Assessment of patient risk; reduction of risky behavior (alcohol consumption) for alcohol-related cancer	RCT	Web-based intervention
Wakefield, 2011[37]	System description and pilot usability test of an online CDS tool that presented 22 men with age and family history-specific prostate cancer risk information and management recommendations. Most participants preferred this method for receiving prostate cancer information	Patients in Australia/New Zealand	FHx	No	Assessment of patient risk and provision of management recommendations for prostate cancer	System description; pilot usability test	Online decision aid using a Markov model

CDS, clinical decision support; FHx, family health history; GPM, genetically guided personalized medicine; RCT, randomized controlled trial.

Table 5 Summary of primary research on CDS systems for pharmacogenomics

Citation and name of system (if applicable)	Manuscript summary and trial details (if applicable)	Users and study location	Genetic information used	Integrated with primary clinical information system	CDS purpose and clinical focus	Manuscript type	Notable informatics aspect
Pazzani, 1997[38]; CTSHIV	System description of the CTSHIV CDS program which manages HIV genome data and makes virus-specific therapeutic recommendations	Clinicians in USA	HIV genotype	No	Provision of management recommendations for HIV	System description	Uses a backward chaining expert system
Tural, 2002[39]; RetroGram	RCT of genotyping accompanied by RetroGram, which ranked drug suitability based on the HIV genotype. This study showed that genotyping combined with RetroGram use improved HIV therapy outcomes over standard of care ($p<0.05$)	Clinicians in Europe	HIV genotype	No	Provision of management recommendations for HIV	RCT	Contains approximately 200 rules based on the scientific literature
Swen, 2008[40]	Description of how the Royal Dutch Association for the Advancement of Pharmacy developed guidelines for the use of genetic information for drug prescribing and integrated these guidelines into automated drug prescription and medical surveillance systems for nationwide use	Clinicians and pharmacists in Europe	Genotype	Yes	Alert on gene-drug interactions for pharmacogenomics	System description	Recommendations incorporated into the G-standard, an electronic drug database used for CDS
Bon Homme, 2008[41]	System description of prototype CDS tool for personalized warfarin therapy that combined genetic and clinical data to estimate the required warfarin dose and the patient's plasma warfarin concentration	Clinicians in USA	Genotype	No	Therapeutic dose guidance for warfarin	System description	Provides a graphical display of estimated plasma warfarin concentration over time
Deshmukh, 2009[42]	This study compared the use of a single nucleotide polymorphism data model to the use of an allele data model for CDS computation in an EHR system. While there were statistically significant differences in computation time, this did not translate into significant differences in the overall clinician ordering time	Clinicians and pharmacists in USA	Genotype	Yes	Alert on gene–drug interactions for pharmacogenomics	Comparative study on genotype data representation	CDS rules developed within the Cerner EHR environment
Overby, 2010[43]	This study found that the Pharmacogenomics Knowledge Base was a good source for pharmacogenomics knowledge and that sufficient clinical data existed in the local EHR system to support 50% of the pharmacogenomic knowledge in drug labels that are capable of being expressed as CDS rules	Clinicians in USA	Genotype	Yes	Provision of therapy guidance for pharmacogenomics	Feasibility study	The MINDscape EHR system was used in the study

CDS, clinical decision support; CTSHIV, Customized Treatment Strategies for HIV; EHR, electronic health record; RCT, randomized controlled trial.

cancer.[47 50 52 53] These systems included a CDS system that retrieved genetic, radiological and clinical data from clinical information systems to provide guidance on intracranial aneurism management,[52] as well as a portable medical device that integrated clinical and genetic data to provide a diagnosis for rheumatoid arthritis and multiple sclerosis.[47] In addition, GeneInsight provides geneticists and other clinicians with patient-specific genetic testing reports, as well as notifications regarding updates to the presumed clinical significance of patients' previously identified genotype.[50] Finally, in a survey study, Scheuner and colleagues[53] found that clinicians felt their EHR systems could do much more to meet their needs related to GPM.

Trend analysis

Publication volume on CDS for GPM generally increased over time, with a majority published since 2007 (figure 3). While all publications before 2007 focused on stand-alone CDS, 32% of articles since 2007 focused on integrated CDS (figure 4). Likewise, while 13% of manuscripts before 2007 involved the use of genotype for CDS, 61% of manuscripts since 2007 have involved the use of genotype (figure 5). As noted earlier, a major focus of the literature in this domain has been on FHx CDS, pharmacogenomics, and CDS for cancer management.

DISCUSSION

Summary of findings

In order to learn from past research efforts and to guide future research into the use of CDS to enable GPM, we conducted a systematic review of the literature. Through a literature search spanning from 1990 to 2011, we screened 3416 manuscripts and included 38 primary research articles. A majority of these manuscripts was published from 2007 to 2011, with an

Table 6 Summary of primary research on GPM CDS systems for other conditions

Citation and name of system (if applicable)	Manuscript summary and trial details (if applicable)	Users and study location	Genetic information used	Integrated with primary clinical information system	CDS purpose and clinical focus	Manuscript type	Notable informatics aspect
FHx-driven CDS systems							
Harris, 1990[44]; GenInfer	System description of the GenInfer program, which used FHx information along with other inheritance factors to calculate genetic risks and probabilities of inheritance	Clinicians in USA	FHx	No	Assessment of patient risk for inherited disease	System description	Based on Pearl's algorithm for fusion and propagation in a probabilistic belief network
Kobrinskii, 1997[45] and Kobrinsky, 1998[46]; National Russian Genetic Register	Description of the information system used by Russia's federal genetics center to manage patients across Russia in need of genetics care. This system supported pedigree creation, cytogenetic analysis, risk assessment, and information support	Genetics specialists in Europe	FHx	No	Assessment of patient risk for inherited disease	System description	Utilized both server–client and local deployment models
Orlando, 2011[48]; MeTree	System description of MeTree, a tool that evaluates FHx and provides management recommendations regarding various heritable conditions for patients and clinicians. Also provides the protocol for a planned evaluation of the tool in North Carolina primary care clinics	Patients and clinicians in USA	FHx	No	Assessment of patient risk and provision of management recommendations for inherited disease	System description; evaluation protocol description	Patient-driven application that provides CDS as a printout
Rubinstein, 2011[49]; CDC Family Healthware	RCT with 3284 participants of the CDC Family Healthware tool, which provides personalized screening recommendations for multiple heritable conditions based on FHx. Both intervention and control groups showed improved adherence to screening recommendations compared to the baseline time period, but there was no significant difference between the intervention and control groups	Patients in USA	FHx	No	Assessment of patient risk and provision of management recommendations for inherited disease	RCT	A patient-directed, web-based tool
Wells, 2007[51]; PREDICT CVD-5	System description of a real-time CDS system that pulled clinical data from the EHR to calculate cardiovascular disease risk and provide risk management recommendations. A retrospective analysis found that including the patients' ethnicity and FHx into the risk assessment process substantially increased the number of patients eligible for drug treatment and lifestyle management	Clinician in Australia and New Zealand	FHx	Yes	Assessment of patient risk for heart disease	System description; retrospective analysis	Integrated with the MedTech practice management system
Genotype-driven CDS systems							
Iavindrasana, 2008[52]; @neurIST	System description of @neurIST, a CDS system which collects genetic data, radiological data, and clinical data from clinical information systems to provide CDS regarding intracranial aneurisms	Clinicians in Europe	Genotype	Yes	Provision of management recommendations for intracranial aneurism	System description	Uses a service- oriented, standards-based approach
Kalatzis, 2009[47]	System description of a point-of-care portable medical device that integrates clinical data with genetic data obtained from a miniature diagnostic system to produce a diagnosis for rheumatoid arthritis and multiple sclerosis	Clinicians in Europe	Genotype	No	Diagnostic assistance for arthritis and multiple sclerosis	System description	A combination of artificial neural networks, decision trees, and support vector machines was found to have the best performance

Continued

Table 6 Continued

Citation and name of system (if applicable)	Manuscript summary and trial details (if applicable)	Users and study location	Genetic information used	Integrated with primary clinical information system	CDS purpose and clinical focus	Manuscript type	Notable informatics aspect
Scheuner, 2009[53]	A survey of health professionals, genetics experts, and EHR developers regarding the ability of EHR systems to document, organize, and use FHx and genetic information	Clinicians in USA	FHx; genotype	Yes	Assessment of patient risk and provision of management recommendations for genetically-guided personalized medicine	Survey	EHRs were generally perceived as lacking the ability to support genomic medicine
Aronson, 2011[50]; GeneInsight	System description of GeneInsight, a platform that provides patient-specific genetic testing reports as well as notifications when the presumed clinical significance of genetic variants change for patients who have been previously tested	Geneticists and other clinicians in USA	Genotype	No	Provision of patient-specific genetic testing reports; notification of changes in clinical significance of genetic variants	System description	Is registered with the Food and Drug Administration as a class I exempt medical device

CDC, Centers for Disease Control and Prevention; CDS, clinical decision support; EHR, electronic health record; FHx, family health history; GPM, genetically guided personalized medicine; RCT, randomized controlled trial.

increasing shift in focus from FHx CDS to genotype-driven CDS, and from stand-alone CDS to integrated CDS. There have been nine RCTs of CDS interventions for GPM, but most CDS interventions for GPM have not yet been rigorously assessed for their clinical impact.

Strengths and limitations

As one important strength of this study, as far as we are aware, this work represents the first systematic review on CDS for GPM. As such, it contributes an important perspective on a topic that has the potential to have significant impacts in both clinical medicine and biomedical informatics. As a second strength, this systematic review was based on search strategies refined through previous systematic reviews on related topics.[14 16] Third, we searched Embase in addition to MEDLINE, so as to provide greater coverage of the international literature. Finally, in addition to providing a summary of relevant manuscripts, this review provides insights and trend analyses that show how this scientific field has developed over time and where the field appears to be headed moving forward.

In terms of limitations, this study does not provide a quantitative meta-analysis of the impact of CDS interventions for GPM. However, such a meta-analysis was not possible due to the limited number of outcome studies in this field and the heterogeneous nature of the various interventions and clinical domains. Second, we only included manuscripts written in English, which may have led to some relevant manuscripts being excluded that were written in a different language. Third, some relevant 2011 articles may not have been indexed by the time of our literature search and therefore erroneously excluded. However, a literature search update in June 2012 added less than 1% to the number of articles we had previously retrieved through March 2012, which suggests that this risk is low. Finally, there is a potential for publication bias with regard to the clinical trials included, in which studies with successful outcomes were more likely to be published than studies with unsuccessful outcomes. There was a potential indication of such a bias, in that seven of nine RCTs evaluated (77%) reported positive results, whereas the expected rate of positive results would more typically be in the range of approximately 60%.[56] However, given the limited sample size, the observed discrepancy may simply be due to chance. Moreover, as discussed next, the high rate of successful interventions may be partly explained by the fact that use of many of these systems was required by the study protocol, which improved the systems' likelihood of use and impact.

Consistency of trial findings with expected outcomes

In a previous systematic review of CDS RCTs, we identified the automatic provision of CDS as a part of routine clinical workflow to be a critical predictor of the success or failure of CDS interventions (adjusted OR of 112.1, $p<0.00001$).[14] While automatic provision of CDS was not a guarantee of success in this systematic review, a lack of this feature was associated with negative outcomes in all cases, generally due to the lack of use of the system.[14] Moreover, a later RCT specifically evaluating the importance of automatic provision of CDS directly confirmed this finding.[57]

On initial examination, the results of the present systematic review seemed to contradict this finding, as we found several RCTs in which stand-alone CDS interventions for GPM were not provided automatically as a part of routine clinical workflow but resulted in positive improvements in clinical practice.[21 25 28 29 36 39] However, in all but one of these RCTs,[21] use of the CDS system

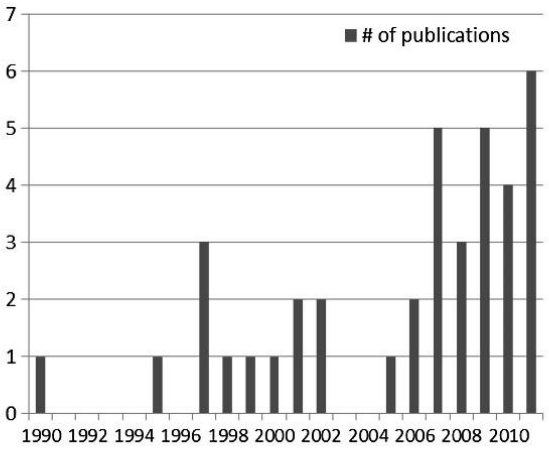


Figure 3 Publications included per year.

was mandated by the study protocol, which was an exclusion criterion in the previous systematic review that identified the critical importance of the automatic provision of CDS.[14] Therefore, we believe it is premature to draw the conclusion that automatic provision is not important when providing CDS for GPM, as it is possible that the same CDS interventions that led to positive results in the studies included would not have led to positive results if use of the system was not mandated by the study protocol, due to lack of awareness and use of the tool. With regard to other, less critical success factors identified in the previous systematic review of CDS interventions,[14] we did not find any trends that contradicted those findings. However, the sample size of available CDS trials was too small in this study to allow for any meaningful analysis of these other factors.

Of note, in the RCT of the GRAIDS system for FHx-based CDS, the system did have a positive impact, even though its use was not mandated by the study protocol and the system was

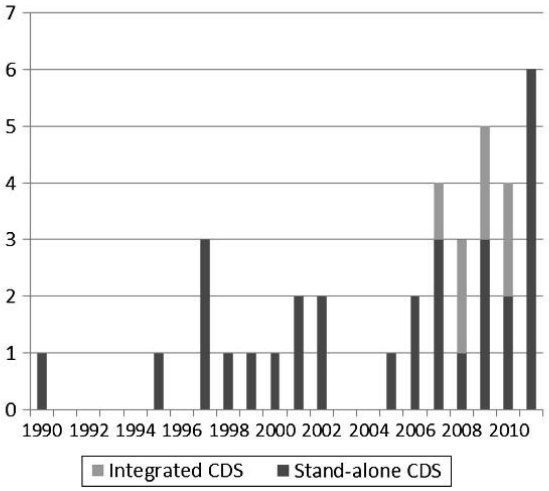


Figure 4 Publications focused on stand-alone versus integrated clinical decision support (CDS).

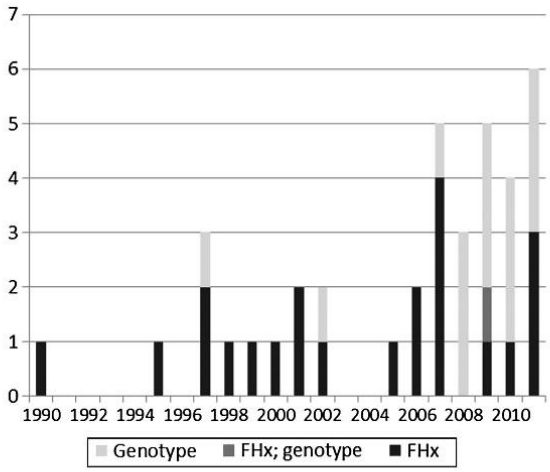


Figure 5 Publications focused on family health history (FHx)-driven versus genotype-driven clinical decision support.

not automatically provided as a part of routine clinical workflow.[14] However, the use and impact of this system may have been the result of exceptional circumstances specific to the study context and unlikely to be available in a routine clinical practice setting. In particular, in the RCT of the GRAIDS system, designated clinicians were recruited at each practice, received extensive training on GRAIDS, and managed all patients in the practice expressing concern regarding their breast or colorectal cancer FHx.[21] This type of resource-intensive deployment strategy may not be feasible outside the context of a research study, as demonstrated in another RCT of a stand-alone breast cancer CDS tool, which had limited impact due largely to the lack of awareness and adoption of the tool by clinicians.[24] Therefore, while more evidence is needed before a solid conclusion can be drawn, we found no conclusive evidence that CDS for GPM is unique in terms of the intervention features required for successful outcomes.

Assessment of current research state and required research

In recent years, CDS has been proposed as a promising approach to realizing the promise of GPM.[10–12 58–65] However, we identified only 38 primary research articles published from 1990 to 2011 on the design, implementation, use, and evaluation of CDS systems to support genetically guided patient care, which amounts to approximately 1.7 articles per year. Even in the year with the most publications on this topic (2011), we identified only six primary research articles. In particular, we identified only nine RCTs of the impact of CDS systems for GPM, seven articles focused on CDS integrated with primary clinical information systems, and 16 articles involving the use of genotype to drive CDS. Furthermore, few groups have demonstrated how genotype-driven CDS can be integrated into clinical settings and clinical information systems in a scalable, standards-based, and effective manner.[40 43 52 53]

Given the tremendous volume of research being conducted in the discovery of novel personalized medicine diagnostics and therapeutics, we feel that much more research is required on how CDS can and should be leveraged to take these discoveries and to implement them in routine clinical practice. For example, even for FHx-driven CDS, which is perhaps the most well-

established area of research with regard to CDS for GPM, there has been limited research on the optimal use of FHx-driven CDS tools beyond hereditary cancer management. Indeed, given the limited literature available on any one topic, we feel it would be premature to consider any aspect of CDS for GPM to be fully mature and not in need of any further research.

In looking forward, we believe that the largest looming research challenge in terms of CDS for GPM will be the development of effective approaches to manage and utilize whole genome sequence data in the clinical setting. The pursuit of low-cost whole genome sequencing has been a priority research area for many years, such that sequencing costs may be reduced to a level amenable to routine clinical use in the near future.[66] While sequencing technologies continue to advance, the informatics capabilities to apply whole genome sequencing data to clinical practice is still in its infancy.[67] Indeed, in our systematic review, we did not find a single primary research article addressing this topic. Therefore, we recommend the prioritization and resourcing of this area of research by the scientific community. In particular, to realize the full clinical potential of whole genome sequence data, we believe that approaches will need to be developed for providing advanced CDS capabilities that are integrated with clinical information systems and provided automatically as a part of routine clinical workflow.

CONCLUSION

The promise of GPM is growing with the recent advances and discoveries in genomics research. With this growth also comes the growing need for translating such discoveries into everyday clinical care, so that we are able to realize the promises of GPM. CDS has the potential to bridge this gap between the promise and realization of GPM. By systematically reviewing the literature in this field and by identifying gaps in required research, we speculate that this paper will assist with efforts to leverage CDS to enable GPM at scale.

Contributors BMW and KK both contributed to the design and conduct of the study, as well as the preparation of the manuscript.

Funding This study was funded by grant K01HG004645 from the US National Human Genome Research Institute, the University of Utah Department of Biomedical Informatics, and the University of Utah Program in Personalized Health Care. The funding sources played no role in the study design, in the collection, analysis and interpretation of data, in the writing of the manuscript, or in the decision to submit the manuscript for publication. The views expressed in this manuscript are those of the authors alone and do not necessarily reflect the official views of the National Human Genome Research Institute or of the University of Utah.

Competing interests BMW is the founder and owner of SGgenomics, Inc., which developed ItRunsInMyFamily.com, a patient-centered FHx tool. KK is serving as a consultant to Inflexxion on a project funded by the National Institute on Drug Abuse to develop CDS capabilities for mental healthcare. KK receives royalties for a Duke University-owned CDS technology for infectious disease management known as CustomID that he helped develop. KK was formerly a consultant for Religent, Inc. and a co-owner and consultant for Clinica Software, Inc., both of which provide commercial CDS services, including through use of a CDS technology known as SEBASTIAN that KK developed. KK no longer has a financial relationship with either Reliqent or Clinica Software.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Individuals interested in the raw data and analyses may contact the corresponding author to obtain such data.

REFERENCES

1. **U.S. Department of Health and Human Services.** *Glossary Of Terms For Personalized Health Care Website*. http://www.hhs.gov/myhealthcare/glossary/glossary.html (accessed Aug 2012).
2. **Abrahams E,** Ginsburg GS, Silver M. The personalized medicine coalition: goals and strategies. *Am J Pharmacogenomics* 2005;**5**:345–55.
3. **Khoury MJ,** McBride CM, Schully SD, *et al*. The scientific foundation for personal genomics: recommendations from a National Institutes of Health–Centers for Disease Control and Prevention multidisciplinary workshop. *Genet Med* 2009;**11**:559–67.
4. **Offit K.** Personalized medicine: new genomics, old lessons. *Hum Genet* 2011;**130**:3–14.
5. **Fernald GH,** Capriotti E, Daneshjou R, *et al*. Bioinformatics challenges for personalized medicine. *Bioinformatics* 2011;**27**:1741–8.
6. **Balas EA,** Boren SA. Managing clinical knowledge for health care improvement. *Yearbook of Medical Informatics 2000: Patient-Centered System*. Stuttgart, Germany: Schattauer, 2000:65–70.
7. **Rosas-Blum E,** Shirsat P, Leiner M. Communicating genetic information: a difficult challenge for future pediatricians. *BMC Med Educ* 2007;**7**:17.
8. **American Board of Genetic Counseling, Inc.** *About ABGC*. http://www.abgc.net/About_ABGC/GeneticCounselors.asp (accessed Aug 2012).
9. *GeneTests*: Growth of Laboratory Directory. Seattle: University of Washington. http://www.ncbi.nlm.nih.gov/projects/GeneTests/static/whatsnew/labdirgrowth.shtml (accessed Aug 2012).
10. **Glaser J,** Henley DE, Downing G, *et al*. Personalized Health Care Workgroup of the American Health Information Community. Advancing personalized health care through health information technology: an update from the American Health Information Community's Personalized Health Care Workgroup. *J Am Med Inform Assoc* 2008;**15**:391–6. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC130085/
11. **Kawamoto K,** Lobach DF, Willard HF, *et al*. A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. *BMC Med Inf Decis Mak* 2009;**9**:17.
12. **Downing GJ,** Boyle SN, Brinner KM, *et al*. Information management to enable personalized medicine: stakeholder roles in building clinical decision support. *BMC Med Inf Decis Mak* 2009;**9**:44.
13. **Osheroff JA,** Teich JM, Middleton B, *et al*. A roadmap for national action on clinical decision support. *J Am Med Inform Assoc* 2007;**14**:141–5.
14. **Kawamoto K,** Houlihan CA, Balas EA, *et al*. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 2005;**330**:765.
15. **Scheuner MT,** Sieverding P, Shekelle PG. Delivery of genomic medicine for common chronic adult diseases: a systematic review. *JAMA* 2008;**299**:1320–34.
16. **Wilson BJ,** Qureshi N, Santaguida P, *et al*. Systematic review: family history in risk assessment for common diseases. *Ann Intern Med* 2009;**151**:878–85.
17. **Coulson AS,** Glasspool DW, Fox J, *et al*. RAGs: a novel approach to computerized genetic risk assessment and decision support from pedigrees. *Methods Inf Med* 2001;**40**:315–22.
18. **Emery J,** Walton R, Coulson A, *et al*. Computer support for recording and interpreting family histories of breast and ovarian cancer in primary care (RAGs): qualitative evaluation with simulated patients. *BMJ* 1999;**319**:32–6.
19. **Emery J,** Walton R, Murphy M, *et al*. Computer support for interpreting family histories of breast and ovarian cancer in primary care: comparative study with simulated cases. *BMJ* 2000;**321**:28–32.
20. **Emery J.** The GRAIDS Trial: the development and evaluation of computer decision support for cancer genetic risk assessment in primary care. *Ann Hum Biol* 2005;**32**:218–27.
21. **Emery J,** Morris H, Goodchild R, *et al*. The GRAIDS Trial: a cluster randomised controlled trial of computer decision support for the management of familial cancer risk in primary care. *Br J Cancer* 2007;**97**:486–93.
22. **Tsouskas LI,** Liaros A, Tzitzikas J, *et al*. The use of an expert system of composite risk factors in breast cancer screening. *Stud Health Technol Inform* 1997;**43**:859–63.
23. **Berry DA,** Iversen ES Jr, Gudbjartsson DF, *et al*. BRCAPRO validation, sensitivity of genetic testing of BRCA1/BRCA2, and prevalence of other breast cancer susceptibility genes. *J Clin Oncol* 2002;**20**:2701–12.
24. **Wilson BJ,** Torrance N, Mollison J, *et al*. Cluster randomized trial of a multifaceted primary care decision-support intervention for inherited breast cancer risk. *Fam Pract* 2006;**23**:537–44.
25. **Matloff ET,** Moyer A, Shannon KM, *et al*. Healthy women with a family history of breast cancer: impact of a tailored genetic counseling intervention on risk perception, knowledge, and menopausal therapy decision making. *J Womens Health* 2006;**15**:843–56.
26. **Matloff ET,** Shannon KM, Moyer A, *et al*. Should menopausal women at increased risk for breast cancer use tamoxifen, raloxifene, or hormone therapy? A framework for personalized risk assessment and counseling. *J Cancer Educ* 2007;**22**:10–14.
27. **Ozanne EM,** Loberg A, Hughes S, *et al*. Identification and management of women at high risk for hereditary breast/ovarian cancer syndrome. *Breast J* 2009;**15**:155–62.
28. **Schwartz MD,** Valdimarsdottir HB, DeMarco TA, *et al*. Randomized trial of a decision aid for BRCA1/BRCA2 mutation carriers: impact on measures of decision making and satisfaction. *Health Psychol* 2009;**28**:11–19.
29. **Hooker GW,** Leventhal KG, DeMarco T, *et al*. Longitudinal changes in patient distress following interactive decision aid use among BRCA1/2 carriers: a randomized trial. *Med Decis Making* 2011;**31**:412–21.
30. **Glasspool DW,** Oettinger A, Smith-Spark JH, *et al*. Supporting medical planning by mitigating cognitive load. *Methods Inf Med* 2007;**46**:636–40.
31. **Glasspool DW,** Oettinger A, Braithwaite D, *et al*. Interactive decision support for risk management: a qualitative evaluation in cancer genetic counselling sessions. *J Cancer Educ* 2010;**25**:312–16.

32. **Evans S,** Lynch HT, Fusaro RM. Clinical results using informatics to evaluate hereditary cancer risk. *Proc Annu Symp Comput Appl Med Care* 1995:834–8.
33. **Bianchi F,** Galizia E, Bracci R, *et al*. Effectiveness of the CRCAPRO program in identifying patients suspected for HNPCC. *Clin Genet* 2007;**71**:158–64.
34. **Overbeek LI,** Hermens RP, van Krieken JH, *et al*. Electronic reminders for pathologists promote recognition of patients at risk for Lynch syndrome: cluster-randomised controlled trial. *Virchows Arch* 2010;**456**:653–9.
35. **Picone M,** Steger S, Exarchos K, *et al*. Enabling heterogeneous data integration and biomedical event prediction through ICT: the test case of cancer reoccurrence. *Adv Exp Med Biol* 2011;**696**:367–75.
36. **Hendershot CS,** Otto JM, Collins SE, *et al*. Evaluation of a brief web-based genetic feedback intervention for reducing alcohol-related health risks associated with ALDH2. *Ann Behav Med* 2010;**40**:77–88.
37. **Wakefield CE,** Watts KJ, Meiser B, *et al*. Development and pilot testing of an online screening decision aid for men with a family history of prostate cancer. *Patient Educ Couns* 2011;**83**:64–72.
38. **Pazzani MJ,** See D, Schroeder E, *et al*. Application of an expert system in the management of HIV-infected patients. *J Acquir Immune Defic Syndr Hum Retrovirol* 1997;**15**:356–62.
39. **Tural C,** Ruiz L, Holtzer C, *et al*. Clinical utility of HIV-1 genotyping and expert advice: the Havana trial. *AIDS* 2002;**16**:209–18.
40. **Swen JJ,** Wilting I, de Goede AL, *et al*. Pharmacogenetics: from bench to byte. *Clin Pharmacol Ther* 2008;**83**:781–7.
41. **Bon Homme M,** Reynolds KK, Valdes R Jr, *et al*. Dynamic pharmacogenetic models in anticoagulation therapy. *Clin Lab Med* 2008;**28**:539–52.
42. **Deshmukh VG,** Hoffman MA, Arnoldi C, *et al*. Efficiency of CYP2C9 genetic test representation for automated pharmacogenetic decision support. *Methods Inf Med* 2009;**48**:282–90.
43. **Overby CL,** Tarczy-Hornoch P, Hoath JI, *et al*. Feasibility of incorporating genomic knowledge into electronic medical records for pharmacogenomic clinical decision support. *BMC Bioinformatics*. 2010;**11**(Suppl. 9):S10.
44. **Harris NL.** Probabilistic belief networks for genetic counseling. *Comput Methods Programs Biomed* 1990;**32**:37–44.
45. **Kobrinskii BA,** Tester IB, Demikova NS, *et al*. A concept of the federal computer system for families with genetic diseases and its practical implementation. *Biomed Eng* 1997;**31**:172–4.
46. **Kobrinsky B,** Tester I, Demikova N, *et al*. A multifunctional system of the national genetic register. *Stud Health Technol Inform* 1998;**52**:121–5.
47. **Kalatzis FG,** Giannakeas N, Exarchos TP, *et al*. Developing a genomic-based point-of-care diagnostic system for rheumatoid arthritis and multiple sclerosis. *Conf Proc Annu Int Conf IEEE Eng Med Biol Soc*; 2–6 September 2009, Minneapolis, Minnesota, **2009**:827–30.
48. **Orlando LA,** Hauser ER, Christianson C, *et al*. Protocol for implementation of family health history collection and decision support into primary care using a computerized family health history system. *BMC Health Serv Res* 2011;**11**:264.
49. **Rubinstein WS,** Acheson LS, O'Neill SM, *et al*. Clinical utility of family history for cancer screening and referral in primary care: a report from the Family Healthware Impact Trial. *Genet Med* 2011;**13**:956–65.
50. **Aronson SJ,** Clark EH, Babb LJ, *et al*. The GeneInsight Suite: a platform to support laboratory and provider use of DNA-based genetic testing. *Hum Mutat* 2011;**32**:532–6.
51. **Wells S,** Kerr A, Broad J, *et al*. The impact of New Zealand CVD risk chart adjustments for family history and ethnicity on eligibility for treatment (PREDICT CVD-5). *NZ Med J* 2007;**120**:U2712.
52. **Iavindrasana J,** Lo Iacono L, Muller H, *et al*. The @neurIST project. *Stud Health Technol Inform* 2008;**138**:161–4.
53. **Scheuner MT,** De Vries H, Kim B, *et al*. Are electronic health records ready for genomic medicine? *Genet Med* 2009;**11**:510–17.
54. **Glasspool DW,** Fox J, Coulson AS, *et al*. Risk assessment in genetics: a semi-quantitative approach. *Stud Health Technol Inform* 2001;**84**:459–63.
55. **Emery J.** Computer support for genetic advice in primary care. *Br J Gen Pract* 1999;**49**:572–5.
56. **Jaspers MW,** Smeulers M, Vermeulen H, *et al*. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc* 2011;**18**:327–34. Epub 2011 Mar 21.
57. **van Wyk JT,** van Wijk MA, Sturkenboom MC, *et al*. Electronic alerts versus on-demand decision support to improve dyslipidemia treatment: a cluster randomized controlled trial. *Circulation* 2008;**117**:371–8.
58. **Ullman-Cullere MH,** Mathew JP. Emerging landscape of genomics in the electronic health record for personalized medicine. *Hum Mutat* 2011;**32**:512–16.
59. **Snyderman R,** Yoediono Z. Prospective care: a personalized, preventative approach to medicine. *Pharmacogenomics* 2006;**7**:5–9.
60. **Samwald M,** Stenzhorn H, Dumontier M, *et al*. Towards an interoperable information infrastructure providing decision support for genomic medicine. *Stud Health Technol Inform* 2011;**169**:165–9.
61. **Hoffman MA.** The genome-enabled electronic medical record. *J Biomed Inform* 2007;**40**:44–6.
62. **Hoffman MA,** Williams MS. Electronic medical records and personalized medicine. *Hum Genet* 2011;**130**:33–9.
63. **Ginsburg GS,** Willard HF. Genomic and personalized medicine: foundations and applications. *Transl Res* 2009;**154**:277–87.
64. **Downing GJ.** Policy perspectives on the emerging pathways of personalized medicine. *Dialogues Clin Neurosci* 2009;**11**:377–87.
65. **Chan IS,** Ginsburg GS. Personalized medicine: progress and promise. *Annu Rev Genomics Hum Genet* 2011;**12**:217–44.
66. **Cordero P,** Ashley EA. Whole-genome sequencing in personalized therapeutics. *Clin Pharmacol Ther* 2012;**2**:51.
67. **Ali-Khan SE,** Daar AS, Shuman C, *et al*. Whole genome scanning: resolving clinical diagnosis and management amidst complex data. *Pediatr Res* 2009;**66**:357–63.

CHAPTER 2

THE NEED FOR CLINICAL DECISION SUPPORT INTEGRATED WITH THE ELECTRONIC HEALTH RECORD FOR THE CLINICAL APPLICATION OF WHOLE GENOME SEQUENCING INFORMATION

Reprinted with permission from:

Welch BM, Kawamoto K. The Need for Clinical Decision Support Integrated with the Electronic Health Record for the Clinical Application of Whole Genome Sequencing Information. *J. Pers. Med*. 2013; 3(4):306-325.

J. Pers. Med. **2013**, *3*, 306-325; doi:10.3390/jpm3040306

OPEN ACCESS

Journal of Personalized Medicine

ISSN 2075-4426

www.mdpi.com/journal/jpm/

Opinion

The Need for Clinical Decision Support Integrated with the Electronic Health Record for the Clinical Application of Whole Genome Sequencing Information

Brandon M. Welch [1,2,*] and Kensaku Kawamoto [2]

[1] Program in Personalized Health Care, University of Utah, 15 North 2030 East, EIHG Room 2110, Salt Lake City, UT 84112, USA

[2] Department of Biomedical Informatics, University of Utah, 26 South 2000 East, Room 5775 HSEB, Salt Lake City, UT 84112, USA; E-Mail: kensaku.kawamoto@utah.edu

* Author to whom correspondence should be addressed; E-Mail: brandon.welch@utah.edu; Tel.: +1-585-455-0461.

Received: 21 October 2013; in revised form: 23 November 2013 / Accepted: 12 December 2013 / Published: 18 December 2013

Abstract: Whole genome sequencing (WGS) is rapidly approaching widespread clinical application. Technology advancements over the past decade, since the first human genome was decoded, have made it feasible to use WGS for clinical care. Future advancements will likely drive down the price to the point wherein WGS is routinely available for care. However, were this to happen today, most of the genetic information available to guide clinical care would go unused due to the complexity of genetics, limited physician proficiency in genetics, and lack of genetics professionals in the clinical workforce. Furthermore, these limitations are unlikely to change in the future. As such, the use of clinical decision support (CDS) to guide genome-guided clinical decision-making is imperative. In this manuscript, we describe the barriers to widespread clinical application of WGS information, describe how CDS can be an important tool for overcoming these barriers, and provide clinical examples of how genome-enabled CDS can be used in the clinical setting.

Keywords: clinical decision support systems; medical genetics; genomics; genetic testing; electronic health records; health information technology; personalized medicine

1. Introduction

Whole genome sequencing (WGS) is on the cusp of revolutionizing medicine. In the decade since the completion of the Human Genome Project, advancements in sequencing technology have made it feasible to sequence a patient's entire genome for clinical uses [1,2]. Indeed, many patients have already had their genome sequenced for direct clinical application to date [3–6]. While most of these cases have been for rare, undiagnosed diseases, or in the context of clinical research, it will not be long before WGS information is available for routine medical care on a widespread scale. This will further enable the practice of personalized medicine, which has the potential to reduce costs and improve the quality of care [7,8].

The WGS information can be used to support clinical diagnosis, direct preventative efforts, and guide therapeutic decisions in the clinic. Indeed, the clinical use of WGS information may hold several advantages over current genetic testing practices:

- There are nearly 3,000 diseases for which individual genetic tests are available [9]. As clinicians pursue a clinical diagnosis today, they sometimes must order several single gene tests until a particular diagnosis is either confirmed or rejected. This process may take a significant amount of time and money as individual genetic tests can cost anywhere between hundreds to thousands of dollars. However, with the ability of WGS to ascertain the results for thousands of available genetic tests at once, it may become financially beneficial and more efficient for clinicians and payers to recommend WGS in lieu of single gene tests, as the diagnostic odyssey and associated costs could be reduced [10].
- Genetic tests are often ordered today as a result of a clinical indication; examples of clinical indications include particular phenotypes, family history, or preliminary diagnosis [11]. This approach is also reinforced by some health insurance providers who require clinical indication and prior authorization in order for certain genetic tests to be reimbursed [12]. However, such an approach can hinder the effective use of genetic information for decision-making, particularly for preemptive and preventative care where clear clinical indications may not always be present [13]. Indeed, if a clinical indication is not present at the time of assessment or clinicians are unaware that a particular genetic test is available, they may miss an opportunity to order the genetic test at a time that can add value to a clinical scenario. Nevertheless, with a patient's WGS information available and readily accessible throughout a patient's life, genetic information can be leveraged for preemptive and preventative care to a larger extent than it is currently.

As WGS is not widely used in the clinical setting at this point, these examples represent theoretical advantages over current genetic testing practices. Until clinical and outcomes research studies on WGS can confirm or reject the validity of these scenarios, these examples will continue to remain theoretical.

Nevertheless, to be most effective, WGS will almost certainly require the effective use of clinical decision support (CDS) integrated into the clinical workflow. However, a systematic review by the authors on the use of CDS for genetically-guided personalized medicine found a significant lack of system descriptions or research studies on the use of CDS to support the clinical use of WGS information [14]. A number of studies identified in the systematic review, as well as several recent

papers and research efforts have described CDS systems that integrate genetics information with CDS. However, these solutions are generally limited in scope with regards to the genetic information used (generally not necessarily WGS), are not integrated within the electronic health record (EHR), or are implemented using CDS approaches that are difficult to scale [15–18]. Indeed, research on CDS solutions for WGS information in particular is still very nascent [18]. While some principles from these efforts can be translated to scalable WGS CDS approaches, additional capabilities will be necessary for the consistent and widespread adoption of CDS capabilities for WGS information [19]. Thus, in order to help establish a foundation for future research and development of scalable CDS for WGS information, this manuscript makes the case for CDS for the WGS. To begin making this case, we start by outlining the many barriers to the effective clinical application of WGS information.

2. Barriers to Effective Clinical Application of WGS Information

Significant barriers exist for the effective and efficient application of WGS information in routine clinical care. These barriers, which will each be described in further detail, include current laboratory reporting methods, the complexity of genetics, the limited physician proficiency in genetics, and the lack of genetics experts. While these barriers have contributed to the slow and inconsistent clinical adoption of genetics [20], we believe that the increased clinical demands as a result of WGS information will make these problems worse.

2.1. Static Laboratory Reports Intended for Human Consumption

Typically, genome sequencing, annotation, and variant classification are performed in Clinical Laboratory Improvement Amendments (CLIA)-approved diagnostic laboratories. If these laboratories follow current standard workflow [21], they will send a static test report by mail, fax, or PDF to the treating clinician. While this workflow has met the needs of current pathology and many genetic tests to date, there are several shortcomings to this approach for WGS information. First, there are roughly three million variants (*i.e.*, mutations) per human genome, and this number is too large to be managed on a single static report. Second, as the genome variant knowledge base continues to grow and change, the need to reclassify variants and notify treating clinicians will become necessary. It is recommended that laboratories take responsibility for updating clinicians to changes in variant interpretation [22]; however, this represents a significant, uncompensated workload upon the laboratory if managed manually. For instance, every individual has hundreds of thousands of variants of unknown significance (VUS), which are variants yet to be associated with a phenotype or ruled out as benign [5]. Over a seven year period, one study found that 14.5% of reported variants had to be reclassified, 27% of which were initially VUS [23]. Third, a static genome report document does not support the automatic provision of CDS at the point of care. Ideally, data should be represented in a discrete, standardized, and digital form accessible to computer interpretation. This is not the case in static laboratory reports. As a result, current laboratory reports require a clinician to manually assess and interpret the reports. At the scale of WGS information, such an approach will likely render most of the information ineffective due to massive information overload [24].

2.2. Complexity of Genetics

Genetics research has brought to light the tremendous complexity of genomic interactions on phenotypes [25]. Within a gene, there can be variants such as point mutations, deletions, insertions, tandem repeats, and splice site mutations, which can all affect the protein product and associated phenotypes. There can be hundreds of possible variants within a specific gene or pathway of genes that contribute to a particular disease etiology. Furthermore, as mentioned previously, not all variants are known to be pathogenic; some are benign while others are VUS [26]. Additionally, gene variants cannot be interpreted in isolation; gene regulatory regions, post-transcriptional modification, transcriptional expression, copy number variations, epistasis, pleiotropy, gene-environment interactions, and other epigenetic influences are additional factors that can modify and impact phenotypes [27]. Relying on a clinician to know all possible genes, variations, and interactions for a particular disease and then to apply that information appropriately at the point of care without assistance is a futile proposition. This is particularly important for common diseases such as heart disease, diabetes, and cancer, which may involve tens to hundreds of contributing genetic, epigenetic, and environmental influences [28]. The interpretation of genetics in the clinic is a complicated endeavor involving numerous genomic interactions and associations which must all be managed accurately for appropriate clinical interpretation.

2.3. Limited Physician Proficiency in Genetics

As stated earlier, there are nearly 3,000 diseases for which an individual genetic test is available [9]. It is beyond the capacity of any human to know and manage all known genetic tests, pertinent genetic contributions, disease-causing variants, and relevant family history associations without computerized support [24,29]. To illustrate, there are almost 1,200 known variants within the adenomatous polyposis coli (*APC*) gene, which is associated with a rare, inherited form of colon cancer [30]. Similarly, almost 2,000 variants in the *CFTR* gene are associated with cystic fibrosis [31]. It is impossible for a clinician to know all possible variants and related variant classifications within a single gene, let alone every variant in the approximately 20,000 genes in the entire human genome. Moreover, genetics is a rapidly growing and evolving field of research; clinicians today do not have the capacity to stay up to date on the current and ever expanding genetics knowledge base [32]. It has been found that it can take 15 or more years for "traditional" medical discoveries to be translated from bench to bedside [33]. Due to the exceeding complexity and breadth of genomics, we expect genetics discoveries to take significantly longer to translate to clinical care with much lower success without additional support.

Furthermore, most clinicians received little to no formal training on the application of genetics to clinical practice [34–36]. Any training they may have received was likely relatively basic, primarily focused on monogenetic diseases with simple inheritance patterns [37,38]. The training required to analyze the complexities of genomics is beyond the scope of most medical school curricula, which are already burdened with numerous competing demands [39]. Accordingly, physicians rate their knowledge of genetics as 'fair to poor' [40,41], with a number of studies confirming their poor knowledge and clinical interpretation of genetics [42,43]. Even education programs specifically designed to teach genetics to clinicians only produce modest results, with substantial gaps in clinician

knowledge on how to appropriately apply genetics at the point of care [43,44]. Expecting clinicians to properly manage a patient's WGS information on their own, even after targeted education, is a daunting proposition.

2.4. Lack of Genetics Professionals

In recent decades, medical genetics and genetic counseling are two specialties that have arisen to provide in-depth knowledge and to help manage the complexity of clinical genetics. Medical geneticists are physicians trained to evaluate patients and assess and manage the genetic contribution to diseases [45]. Genetic counselors are masters-level health professionals who work with physicians to help assess genetic risk and communicate genetic information, such as test results, to patients and their families [46]. While these specialties attempt to fill the need, there are only 1,200 medical geneticists and 3,000 certified genetic counselors who are unequally and insufficiently distributed across the United States today [47,48]. Today, a genetic professional typically spends seven hours preparing for and meeting with new patients [49]. With only one medical geneticist per 262,000 U.S. citizens and one genetic counselor per 105,000 U.S. citizens, each genetic professional would have to work for over 239 years if every person in the U.S. had their genome sequenced today [50]. The insufficient supply of genetics experts is unlikely to change in the foreseeable future, as aspiring clinicians are not entering the genetics profession at the rate needed for growth [51]. Furthermore, with the potential for genetics to impact so many clinical decisions [52], and the labor intensive nature of genetic interpretation and counseling [49], it will be inadequate, inefficient, and cost-prohibitive to have a genetics professional available every time genetic information is used at the point of care [53].

3. CDS as a Solution

Clearly, significant barriers exist which will hinder the effective and efficient application of genetics at the point of care. Nevertheless, a number of thought leaders and researchers have recognized this problem and have identified EHRs incorporating CDS as a practicable solution to help clinicians manage the complexities of genetics at the point of care [53–57]. CDS entails providing clinicians, patients, and other healthcare stakeholders with pertinent knowledge and/or person-specific information, intelligently filtered or presented at appropriate times, to enhance health and healthcare [58]. Examples of CDS include medication dosing support, order facilitators, point of care alerts and reminders, relevant information display, expert systems, and workflow support [59]. Research on CDS has been conducted for decades and is a proven solution for assisting clinicians in providing appropriate care and reducing errors in many clinical use cases [60–63]. Furthermore, CDS has the ability to translate knowledge from bench to bedside much more efficiently than traditional methods [54]. As a result of this success, the Office of the National Coordinator for Health IT (ONC) has announced that CDS will be a key component of proposed Meaningful Use Stage 3 criteria, which are expected to be proposed in 2014 [64].

3.1. Overcoming WGS Barriers

CDS has been identified as a potential solution for supporting WGS information in the clinic because of its ability to overcome many of the barriers to effective use of WGS information described above. CDS has the capacity to process complex, disparate clinical data and present actionable, evidence-based recommendations in a way that is usable by a clinician at the point of care [63]. This capability will be essential for CDS for WGS information because certain clinical use cases may require the combination of several genetic loci as well as the patient's health history, family history, and environmental influences to develop accurate clinical assessments and recommendations. Furthermore, CDS is able to automate the application of complex decision logic and provide clinically actionable information to the treating clinician. As such, CDS allows clinicians to focus on caring for patients rather than interpreting complex WGS information, for which they are not traditionally trained to do. Likewise, CDS that provides clear, clinically actionable recommendations derived from WGS information will allow clinicians, even those with minimal training in genetics, to harness WGS information and improve the care of their patients. Moreover, when widespread use of WGS has outpaced the capacity of available genetics professionals, CDS will be able to meet such demands on a widespread scale [19]. This is not to say that genetic professionals will be replaced by CDS, but rather that CDS can manage the common, routine applications of WGS information while allowing genetics professionals to focus their expertise and effort on novel clinical applications of WGS information.

3.2. CDS Best Practices

While CDS offers a potential solution to overcome the clinical barriers to WGS adoptions, it is important to consider that not all CDS interventions are successful. Indeed, a systematic review showed that CDS interventions only improved clinical performance about two-thirds of the time [65]. Through practical experience and systematic reviews, researchers have identified important features that contribute to a CDS system being effective. Bates *et al.* summarized their experiences implementing CDS with 'Ten Commandments for Effective Clinical Decision Support,' summarized in Table 1 [66]. Kawamoto *et al.* found that computer-generated CDS interventions which are provided automatically during clinical workflow, at the time and location of care, and as care recommendations rather than assessments, are successful more than 90% of the time. This same study found that if any of those key factors were missing, the CDS intervention was successful less than 50% of the time [67]. When developing CDS for WGS, it will be necessary to adhere to these best practices so that the developed CDS solutions have the greatest chance of being successful. Indeed, if such features are not incorporated into CDS interventions for WGS, it will run the risk of failing simply because basic best practices for CDS implementations are not adhered to.

3.3. CDS for WGS

CDS for WGS will need to integrate into the clinical workflow and seamlessly provide support at the location and time of decision making, in a manner consistent with best practices. Ideally, CDS would be provided automatically at the time of care within the workflow of the clinician's EHR or other primary health information systems, such as computerized provider order entry (CPOE) systems.

Indeed, clinicians should not be required to use a separate CDS application for WGS information nor even have to initiate the CDS within their primary clinical information system [68]. Moreover, the automatic CDS recommendations could be provided in such a way that the clinician end user would not even need to know that WGS information was being used to generate the recommendation. Preferably, CDS capabilities developed for WGS would not be separate from non-WGS CDS. Rather, WGS information should just be another source of information available for CDS.

Table 1. Ten Commandments for effective clinical decision support by Bates *et al.* [66].

1.	Speed is Everything
2.	Anticipate Needs and Deliver in Real Time
3.	Fit into the User's Workflow
4.	Little Things Can Make a Big Difference
5.	Recognize that Physicians Will Strongly Resist Stopping
6.	Changing Direction is Easier than Stopping
7.	Simple Interventions Work Best
8.	Ask for Additional Information Only When You Really Need It
9.	Monitor Impact, Get Feedback, and Respond
10.	Manage and Maintain Your Knowledge-based Systems

4. Potential Clinical Applications of CDS for WGS Information

Genetic information can have many applications in health care. Here, to further make the case for CDS for WGS information, we describe several clinical use cases in which genomic information can be used to guide care. In each example, we propose how automatic CDS leveraging WGS information might be integrated within the clinical workflow. While this is not a comprehensive list, it illustrates various examples of genome-enabled CDS applications at the point of care.

4.1. Clinical Diagnosis

Genetic testing is traditionally used to confirm or rule out a diagnosis during the differential diagnosis process [69]. Often, however, clinicians may not know a relevant genetic test is available to support their decision making process. Indeed, some patients often wait months to years to receive an accurate diagnosis, even after seeing several specialists [70]. With a patient's WGS information readily available, an accurate diagnosis can be reached faster [71,72]. However, for this to happen, it is important to make relevant genetic information easily accessible and reviewable to clinicians at the point of care. While working up a diagnosis, a clinician may not know or be aware of all known genes associated with particular symptoms. At the very least, clinicians should have a simple list of genes containing known pathogenic or likely-pathogenic variants, disease names, and associated phenotypes which the clinician can refer to during the differential diagnosis process. Based upon our current understanding of the human genome, the list of clinically relevant variants will be relatively short (hundreds) [73]. Thus, clinicians could review this information and match it to the patient's phenotypes; the list should also have the option to view genes containing VUS. Furthermore, disease-causing variants could be automatically added to the EHR's problem list.

Medical geneticists often search a genetics knowledge base like Online Mendelian Inheritance in Man (OMIM) using phenotypes to identify potentially disease-causing genes for which they could order a genetic test to assess the genotype [74]. Ideally, with the patient's genome readily available, CDS capabilities integrated with the EHR could automatically perform an 'OMIM-like' search on the patient's genome for candidate gene variants based on phenotypes documented in the EHR. For example, for a child presenting at a pediatric clinic with deafness and heterochromia, which is recorded in the EHR problem list, a CDS system could automatically search the patient's genome (assuming it is already available) for potential pathogenic variants in genes associated with the presented phenotypes. In this case, if a pathogenic mutation was identified in the patient's *PAX3* gene, a recommendation can be provided to add Waardenburg Syndrome, a rare genetic disorder, to the problem list [75]. Furthermore, the CDS system could generate a referral to a clinical expert specializing in hereditary deafness who is also covered by the patient's insurance. While these capabilities could be available in stand-alone genome management products, to be clinically effective they must be integrated within the EHR and clinical workflow.

4.2. Disease Risk Assessment

Diseases can be associated with a number of risk factors. Thus, using genetic testing to estimate disease risk is an important aspect of predictive medicine. A gene variant's influence on disease can range from a slight increase in disease risk to a certainty of future disease onset. With genetic information available, the risk for certain diseases can be deduced, which can then lead to preventative and risk-reducing actions for the patient. To illustrate, women with mutations in the *BRCA1* or *BRCA2* genes have a 50%–80% chance of developing breast cancer in their lifetimes [76]. Knowing this information beforehand can allow women to take risk-reducing actions such as increased screening and prophylactic mastectomy. Unfortunately, as a result of the shortcomings of single gene tests [13], it is estimated that only 5% of women with *BRCA* mutations have been identified with genetic testing [77].

With WGS information readily available, a genome-enabled CDS system could systematically assess the patient's WGS and clinical information and provide disease risk estimations and risk-reducing recommendations. Such a capability would alleviate the need for clinicians to estimate disease risk on their own. For example, the presence of a pathogenic variant in the *BRCA1* gene (or other genes associated with breast cancer) in a pre-menopausal woman not desiring to have a prophylactic mastectomy could trigger a CDS system to pre-populate an order for more frequent mammograms and to prescribe a selective estrogen receptor modulator, such as tamoxifen. Also, as described above, elevated risk for a disease could automatically be populated on the EHRs problem list. There are a number of stand-alone CDS solutions that provide risk assessment and recommendations for *BRCA* gene mutations [78,79]. However, an ideal scenario would be for such solutions to automatically leverage WGS information and be tightly integrated with the EHR so recommendations are provided within the clinical workflow.

4.3. Reproductive Carrier Screening

Related to one's own disease risk assessment is reproductive carrier screening. Hundreds of congenital disorders are caused by the inheritance of gene variants by one or both parents. Carrier

screening is a genetic test that can identify the presence of a disease-causing genetic variant in one or both parents. With this knowledge, it is possible to estimate the risk of having a child affected with a genetic disease, allowing parents to make more informed reproductive decisions. For example, cystic fibrosis is a recessive newborn genetic disorder affecting one in 3,500 births in the U.S. [80]. If both parents are discovered to be carriers for *CFTR* mutations, they can opt for adoption or preimplantation genetic diagnosis to reduce their chances of having a child affected with cystic fibrosis.

Again, CDS could provide support by automatically assessing the patient's genome to assess their genetic carrier status, and referring the patient to a reproductive specialist or genetic counselor if he or she is considering reproduction. Of note, with autosomal recessive genetic diseases, one parent's genome is typically only 'half the equation'; both parental genomes are required to accurately predict disease risks. One could therefore envision an inheritance risk assessment CDS application within the EHR that is able to access *both* parental genomes and provide recommendations only when a risk of having an affected child is present. Similar features are available from commercial laboratory prenatal genetic testing companies such as Counsyl [81]. However, with WGS information readily available for point-of-care CDS, such capabilities could be managed directly by CDS and results presented within the EHR, without the need for another genetic test.

4.4. Pharmacogenomics

After a diagnosis has been made, genomic information can be used to guide appropriate therapy and accurate drug dosing. A commonly used example for pharmacogenomics is the use of the anticoagulant warfarin (Coumadin) to prevent thrombosis. Warfarin is metabolized by enzymes derived from the *VKORC1* and *CYP2C9* genes [82]. Variants in these genes can cause patients to be rapid metabolizers of the drug, thereby causing standard dosing regimens to be ineffective. Alternatively, variants can cause patients to be slow metabolizers of the drug, resulting in the drug persisting in the blood longer than expected and accumulating to toxic levels when standard therapeutic doses are administered. By assessing for such gene variants prior to drug therapy, the clinician can reach an optimal therapeutic dose faster while also avoiding adverse events or ineffective treatment regimens, saving lives and unnecessary costs.

Already widely deployed within clinical workflows are drug-drug, drug-allergy, and drug-condition interaction checking within CPOE systems to prevent adverse drug events. These existing CDS capabilities could potentially be extended with WGS information to support drug-gene interaction checking. When a drug like warfarin is prescribed, in addition to checking patient-specific information such as age, weight, and other medications, the CDS rule could also automatically assess the patient's genome for variants in related genes, such as *VKORC1* and *CYP2C9*, and alert the ordering clinician to any potential complications. Ideally, such a CDS system will not just check for adverse interactions but also provide an optimized drug dose recommendation to the prescribing clinician based on available genetic information and pertinent clinical information [83]. For CDS to be most effective in guiding pharmacogenomics, it needs to be provided within the workflow of the clinician, ideally at the point of order entry. It cannot be expected that a clinician will manually review a patient's genome for relevant genes variants and then use a stand-alone CDS application, like www.WarfarinDosing.org, every time a drug is prescribed [84]. Some EHR-integrated pharmacogenomics CDS capabilities have been

implemented and investigated by researchers at Vanderbilt and St. Jude Children's Hospital [17,85]. A future step in this research would be to enable the CDS to leverage WGS information and to use CDS capabilities that can be scaled to other institutions.

4.5. Nutritional Genomics

The same enzymes that are involved with drug metabolism are also involved in nutrient metabolism. In fact, these enzymes originally evolved for diet; it is only recently in our evolutionary history that clinical medicine has leveraged these enzymes for therapeutics. As such, like pharmacogenomics, nutritional considerations can be personalized based on genetics information. Given that nutrition impacts health, it is essential that clinicians manage nutritional variability caused by genetics. For example, choline, an essential nutrient, is known to be affected by a common variant in the *MTHFD1* gene [86]. People with this variant, particularly pregnant women, need to eat foods rich in choline to avoid adverse consequences of choline deficiency such as neural tube defects in unborn fetuses [87]. Like previous examples, genome-enabled CDS can assess a patient's genome for such variants and notify the clinician and/or the patient of this risk and make recommendations on how to maintain a diet high in choline.

5. Future Direction

A 2012 systematic review [14], as well as more recent manuscripts published after the search period covered by the review describe CDS systems that leverage genetic information [17,18]. With only a few exceptions, these systems primarily leverage a single or a few genes and are typically not automatically integrated within the clinical workflow of the EHR. Nevertheless, these examples represent necessary and important steps toward an ideal CDS solution for WGS information. However, for CDS capabilities to fully meet the demands of WGS information, a number of challenges must be addressed. As such, these challenges will require new CDS approaches that are able to support the unique demands of WGS information. Indeed, a significant amount of work need to be done before WGS information is efficiently incorporated into busy clinical settings through CDS [88,89].

5.1. Challenges to Overcome

Several challenges will need to be addressed for the vision of CDS for WGS information to be realized. For example, our understanding of the human genome and its implication on health is still relatively nascent. Research into the genetic contribution to disease has been limited by our ability to leverage a sufficient amount of genetic information and high-quality phenotypic information [90]. The combination of falling genome sequencing costs and EHRs becoming better at representing structured phenotypic information will improve researchers' ability to identify disease-causing genetic variants on a large scale [91]. These new discoveries are anticipated to lead to many important, clinically relevant recommendations that can be used to improve clinical care. Furthermore, as the genomics knowledge base continues to grow, recommendations will continue to change and evolve with new knowledge. Thus, it is important that CDS be a conduit through which such discoveries can be efficiently translated into clinical care on a widespread scale.

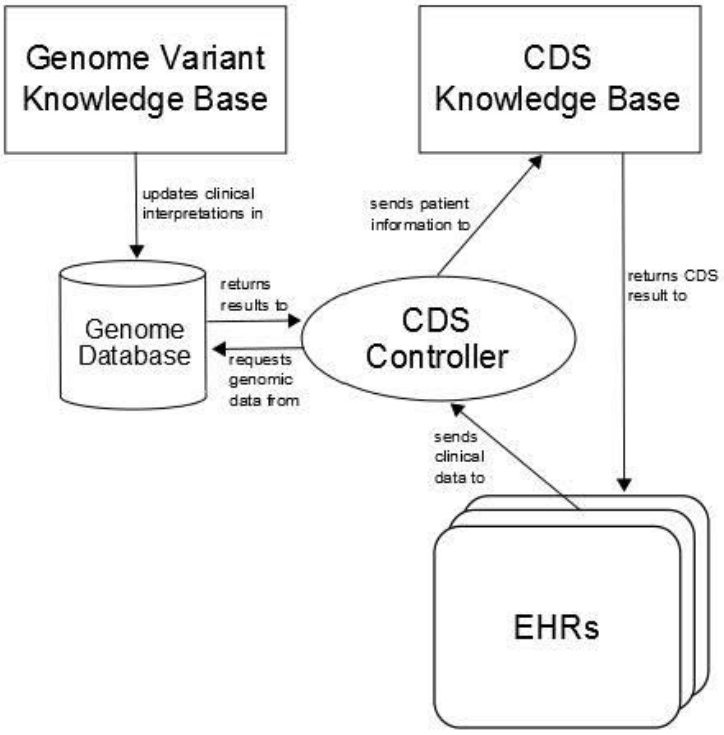
In addition to the changing clinical knowledge base, an important challenge to WGS-based CDS is the need to maintain a constantly growing and evolving genome variant knowledge base [23]. As described earlier, every patient has many variants in genes associated with diseases. Currently, it is the task of clinical genetic testing laboratories to assign a clinical interpretation (e.g., pathogenic, benign) to the variants detected. Often, however, many interpretations are variants of unknown significance (VUS) because not enough information is known about the variant identified. Although it is the responsibility of the laboratory to notify clinicians to changes in variant interpretation, this can become a daunting and uncompensated task for laboratories to manage because a significant portion of variant interpretations will need to be changed over time. Furthermore, testing laboratories often use their own proprietary repository of variant interpretations and/or many independently-managed gene or disease-specific variant knowledge bases to help make variant interpretations. These variant knowledge bases may only represent a subset of all known variants. Indeed, a centrally-managed gene variant knowledge base would provide great value to improving genome variant interpretation. It is important to note that two large federally-funded efforts, ClinVar and ClinGen, seek to create large publically-available gene and variant knowledge bases to help facilitate gene variant interpretation in the future [92]. However, as these initiatives have just started, it may still be a few years before the full potential of these resources are realized for CDS [93].

Beyond the challenges related to the genome are challenges related to health IT infrastructures that CDS capabilities will be dependent upon. Traditionally, it has been difficult to integrate EHR systems with most third-party CDS applications [94]. While Meaningful Use requirements are making it possible to do more with EHRs, these systems are still challenging to change or to integrate external applications. Therefore, it will be important for CDS capabilities for WGS information to leverage the available features of EHRs, many of which are being made available as a result of Meaningful Use. Furthermore, many CDS capabilities currently available within EHRs are not scalable beyond the institutions at which they are created [95]. Given the time and effort needed to create CDS rules and the potentially extensive list of CDS interventions that could leverage WGS information, the limited scalability of CDS solutions currently available in EHRs will not be sufficient for meeting the full potential of WGS-based CDS. Indeed, current efforts are underway to develop Meaningful Use requirements for EHRs to support service-based CDS capabilities [96]. Without these new CDS service capabilities enabled for EHRs, it will be challenging to support CDS for the WGS on a widespread scale with current CDS capabilities [62,97].

5.2. Proposed Solution

Given the possibility of leveraging service-based CDS capabilities enabled by Meaningful Use, a standards-based and scalable CDS solution could be developed for WGS information [19]. Given the breadth and complexity of the human genome and the rapidly growing knowledge base, it is unlikely that any single EHR vendor or health care organization will be able to fully manage genomic capabilities on its own. Thus, achieving effective CDS for WGS information will likely require the coordination of several independent services or entities managing specific tasks. See Figure 1. Independent services components that would need to be coordinated could include the EHR, a genome database, variant knowledge base, a CDS knowledge base, and a CDS controller.

Figure 1. A graphical representation of a proposed scalable clinical decision support (CDS) architecture that can leverage whole genome sequencing (WGS) information.



In such a scenario, when a CDS service request for WGS information is initiated by an EHR, the request with standardized, structured clinical information document, such as the Health Level 7 Virtual Medical Record (vMR) or Consolidated Clinical Document Architecture (C-CDA), could be sent to the CDS controller which parses the received clinical document into the data required for the rule evaluation that is being requested. The CDS controller would also identify which genetic information is required and submit a query to a genome database for the patient's genetic information and variant interpretation at a particular loci. The genome database, which stores patients' genetic information and interpretations, could have its variant interpretations updated and maintained by a separate genome variant knowledge base such as ClinVar. With the most up-to-date clinical interpretation and genetic information returned to the CDS controller, the full set of patient information needed for the CDS evaluation can be sent to and processed by the CDS knowledge base. The WGS-enabled CDS result is then returned to the EHR for presentation within the clinical workflow at the point and time of care.

Such a solution could allow CDS for WGS to be implemented on a widespread scale with the most accurate and up-to-date information available at any given time. Moreover, as rules could be developed and used by many organizations, the economies of scale to implement new CDS recommendations for WGS information is much lower than if each organization attempted to develop and deploy the same capabilities on their own. Importantly, this approach is aligned with current and future EHR capabilities to support service-based CDS, as anticipated by Meaningful Use Stage 3

requirements. While this infrastructure is largely theoretical construct at this point, efforts are currently underway to develop and validate this proposed CDS architecture approach for WGS information. Nevertheless, before such as architecture can be implemented in a clinical setting on a widespread scale, several considerations need to be resolved. For example: What genetic information is sufficient and necessary for CDS? Who manages and controls each component of the proposed architecture? How can several independent components be coordinated to promote efficiency? Furthermore, it should also be noted that it is yet to be determined if this proposed architecture can meet all needs of WGS information. Indeed, there may be scenarios where a different architecture would be better for certain use cases. This will be determined through continued research and development on CDS capabilities for WGS information.

Finally, clinicians and health IT vendors need to be aware of the coming deluge of genomic information to the clinic so they can be prepared to respond accordingly. Researchers already grapple with overwhelming amounts of genetic information. It is well known by genetics experts that the WGS technologies currently being used in research settings will soon be available to everyday clinicians. Health care organizations and health IT vendors need to be proactive in developing clinical information systems that have the capacity to leverage the WGS in an effective manner. It will likely be impossible for any single health care organization or health IT vendor to manage and support all aspects of genome interpretation and CDS capabilities. As such, health care organizations may need to leverage third party CDS providers, and EHRs need the capacity to support distributed computing architectures, like the one just described, that are capable of integrating external CDS capabilities.

6. Conclusions

We anticipate that WGS capabilities will eventually be routinely available for clinical care. However, without appropriate support, WGS information will likely overwhelm clinicians because of current laboratory reporting methods, the complexity of genetic information, the limited proficiency in genetics by most physicians, and the lack of genetics professionals. However, CDS capabilities can overcome these barriers and increase the likelihood that WGS information can be used effectively for clinical care. Nevertheless, it will be essential that CDS be provided within the clinical workflow and at the point of care in the EHR according to established CDS best practices. We described several clinical use cases using WGS information and described how CDS could be provided within the EHR to support these clinical use cases. Key next steps will be to design and develop a scalable CDS framework capable of leveraging complex WGS information on a widespread scale.

Acknowledgments

We would like to acknowledge the University of Utah Program in Personalized Health Care and the University of Utah, Department of Biomedical Informatics for sponsoring this research.

Conflicts of Interest

KK is currently or recently served as a consultant on CDS to the Office of the National Coordinator for Health IT (Washington, DC, USA), McKesson InterQual (San Francisco, CA, USA), ESAC, Inc.

(Enterprise Science and Computing; Rockville, MD, USA), ARUP Laboratories (Associated Regional and University Pathologists; Salt Lake City, UT, USA), Inflexxion, Inc. (Newton, MA, USA), Intelligent Automation, Inc. (Rockville, MD, USA), Partners HealthCare (Boston, MA, USA), and the RAND Corporation (Research and Development; Santa Monica, CA, USA). KK receives royalties for a Duke University-owned CDS technology for infectious disease management known as CustomID (Durham, NC, USA) that he helped develop. KK was formerly a consultant for Religent, Inc. (Raleigh, NC, USA) and a co-owner and consultant for Clinica Software, Inc. (Durham, NC, USA), both of which provide commercial CDS services, including through use of a CDS technology known as SEBASTIAN (System for Evidence-Based Advice through Simultaneous Transaction with an Intelligent Agent across a Network; Durham, NC, USA) that KK developed. KK no longer has a financial relationship with either Religent or Clinica Software.

References

1. Wetterstrand, K. DNA sequencing costs: Data from the NHGRI Genome Sequencing Program (GSP). Available online: http://www.genome.gov/sequencingcosts/ (accessed on 6 February 2013).
2. Bonetta, L. Whole-genome sequencing breaks the cost barrier. *Cell* **2010**, *141*, 917–919.
3. Rope, A.F.; Wang, K.; Evjenth, R.; Xing, J.; Johnston, J.J.; Swensen, J.J.; Johnson, W.E.; Moore, B.; Huff, C.D.; Bird, L.M.; *et al.* Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to *N*-terminal acetyltransferase deficiency. *Am. J. Hum. Genet.* **2011**, *89*, 28–43.
4. Ashley, E.A.; Butte, A.J.; Wheeler, M.T.; Chen, R.; Klein, T.E.; Dewey, F.E.; Dudley, J.T.; Ormond, K.E.; Pavlovic, A.; Morgan, A.A.; *et al.* Clinical assessment incorporating a personal genome. *Lancet* **2010**, *375*, 1525–1535.
5. Lupski, J.R.; Reid, J.G.; Gonzaga-Jauregui, C.; Rio Deiros, D.; Chen, D.C.; Nazareth, L.; Bainbridge, M.; Dinh, H.; Jing, C.; Wheeler, D.A.; *et al.* Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N. Engl. J. Med.* **2010**, *362*, 1181–1191.
6. Talkowski, M.E.; Ordulu, Z.; Pillalamarri, V.; Benson, C.B.; Blumenthal, I.; Connolly, S.; Hanscom, C.; Hussain, N.; Pereira, S.; Picker, J.; *et al.* Clinical diagnosis by whole-genome sequencing of a prenatal sample. *N. Engl. J. Med.* **2012**, *367*, 2226–2232.
7. Report of the President's Council of Advisors on Science and Technology (PCAST). Priorities for Personalized Medicine. Available online: http://www.whitehouse.gov/files/documents/ostp/PCAST/ pcast_report_v2.pdf (accessed on 19 February 2013).
8. Abrahams, E.; Ginsburg, G.S.; Silver, M. The Personalized medicine coalition: Goals and strategies. *Am. J. Pharmacogenomics* **2005**, *5*, 345–355.
9. GeneTests Medical Genetics Information Resource (database online). Available online: http://www.genetests.org/ (accessed on 6 February 2013).
10. McGowan, K. Genomic Information Wants to Be Free, Says Randy Scott at PMWC. Available online: http://nygenome.org/blog/genomic-information-wants-be-free-says-randy-scott-pmwc/ (accessed on 8 February 2013).
11. Heart Rhythm UK Familial Sudden Death Syndromes Statement Development Group. Clinical indications for genetic testing in familial sudden cardiac death syndromes: An HRUK position statement. *Heart* **2008**, *94*, 502–507.

12. Carlson, B. Payers try new approaches to manage molecular diagnostics. *Biotechnol. Healthc.* **2010**, *7*, 26–30.
13. Weldon, C.B.; Trosman, J.R.; Gradishar, W.J.; Benson, A.B.; Schink, J.C. Barriers to the use of personalized medicine in breast cancer. *J. Oncol. Pract.* **2012**, *8*, e24–e31.
14. Welch, B.M.; Kawamoto, K. Clinical decision support for genetically guided personalized medicine: A systematic review. *J. Am. Med. Inform. Assoc.* **2012**, *20*, 388–400.
15. Bon Homme, M.; Reynolds, K.K.; Valdes, R., Jr.; Linder, M.W. Dynamic pharmacogenetic models in anticoagulation therapy. *Clin. Lab. Med.* **2008**, *28*, 539–552.
16. Glasspool, D.W.; Oettinger, A.; Braithwaite, D.; Fox, J. Interactive decision support for risk management: A qualitative evaluation in cancer genetic counselling sessions. *J. Cancer Educ.* **2010**, *25*, 312–316.
17. Bell, G.C.; Crews, K.R.; Wilkinson, M.R.; Haidar, C.E.; Hicks, J.K.; Baker, D.K.; Kornegay, N.M.; Yang, W.; Cross, S.J.; Howard, S.C.; *et al.* Development and use of active clinical decision support for preemptive pharmacogenomics. *J. Am. Med. Inform. Assoc.* **2013**, doi:10.1136/amiajnl-2013-001993.
18. Tarczy-Hornoch, P.; Amendola, L.; Aronson, S.J.; Garraway, L.; Gray, S.; Grundmeier, R.W.; Hindorff, L.A.; Jarvik, G.; Karavite, D.; Lebo, M.; *et al.* A survey of informatics approaches to whole-exome and whole-genome clinical reporting in the electronic health record. *Genet. Med.* **2013**, *15*, 824–832.
19. Kawamoto, K.; Lobach, D.F.; Willard, H.F.; Ginsburg, G.S. A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. *BMC Med. Inform. Decis. Mak.* **2009**, *9*, e17.
20. Hamilton, A.B.; Oishi, S.; Yano, E.M.; Gammage, C.E.; Marshall, N.J.; Scheuner, M.T. Factors influencing organizational adoption and implementation of clinical genetic services. *Genet. Med.* **2013**, doi:10.1038/gim.2013.101.
21. Green, R.C.; Rehm, H.L.; Kohane, I.S. Clinical genome sequencing. In *Genomic and Personalized Medicine*, 2nd ed.; Academic Press: London, UK, 2013; pp. 102–122.
22. Richards, C.S.; Bale, S.; Bellissimo, D.B.; Das, S.; Grody, W.W.; Hegde, M.R.; Lyon, E.; Ward, B.E. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet. Med.* **2008**, *10*, 294–300.
23. Aronson, S.J.; Clark, E.H.; Varugheese, M.; Baxter, S.; Babb, L.J.; Rehm, H.L. Communicating new knowledge on previously reported genetic variants. *Genet. Med.* **2012**, *14*, 713–719.
24. Masys, D.R. Effects of current and future information technologies on the health care workforce. *Health Aff.* **2002**, *21*, 33–41.
25. West, M.; Ginsburg, G.S.; Huang, A.T.; Nevins, J.R. Embracing the complexity of genomic data for personalized medicine. *Genome Res.* **2006**, *16*, 559–566.
26. Domchek, S.; Weber, B.L. Genetic variants of uncertain significance: Flies in the ointment. *J. Clin. Oncol.* **2008**, *26*, 16–17.
27. Jorde, L.B.; Carey, J.C.; Bamshad, M.J. *Medical Genetics*, 4th ed.; Mosby: Maryland Heights, MO, USA, 2009; p. 368.
28. Scheuner, M.T.; Sieverding, P.; Shekelle, P.G. Delivery of genomic medicine for common chronic adult diseases: A systematic review. *JAMA* **2008**, *299*, 1320–1334.

29. Cowan, N. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav. Brain Sci.* **2001**, *24*, 87–114.
30. Colon cancer gene variant databases. Adenomatous Polyposis Coli (APC) Available online: http://chromium.liacs.nl/LOVD2/colon_cancer/home.php?select_db=APC/ (accessed on 19 February 2013).
31. Cystic Fibrosis Mutation Database: Statistics. Available online: http://www.genet.sickkids.on.ca/StatisticsPage.html/ (accessed on 19 February 2013).
32. Shirts, B.H.; Parker, L.S. Changing interpretations, stable genes: Responsibilities of patients, professionals, and policy makers in the clinical interpretation of complex genetic information. *Genet. Med.* **2008**, *10*, 778–783.
33. Balas, E.A.; Boren, S.A. Managing clinical knowledge for health care improvement. In *Yearbook of Medical Informatics*; Bemmel, J., McCray, A.T., Eds.; Patient-Centered Systems: Stuttgart, Germany, 2000; pp. 65–70.
34. Thurston, V.C.; Wales, P.S.; Bell, M.A.; Torbeck, L.; Brokaw, J.J. The current status of medical genetics instruction in US and Canadian medical schools. *Acad. Med.* **2007**, *82*, 441–445.
35. Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS). *Genetics Education and Training of Health Care Professionals, Public Health Providers, and Consumers Service*; SACGHS: Bethesda, MD, USA, 2010.
36. Collins, F.S.; Bochm, K. Avoiding casualties in the genetic revolution: The urgent need to educate physicians about genetics. *Acad. Med.* **1999**, *74*, 48–49.
37. Greb, A.E.; Brennan, S.; McParlane, L.; Page, R.; Bridge, P.D. Retention of medical genetics knowledge and skills by medical students. *Genet. Med.* **2009**, *11*, 365–370.
38. McInerney, J.D.; Edelman, E.; Nissen, T.; Reed, K.; Scott, J.A. Preparing health professionals for individualized medicine. *Pers. Med.* **2012**, *9*, 529–537.
39. McInerney, J.D. Genetics education for health professionals: A context. *J. Genet. Couns.* **2008**, *17*, 145–151.
40. Hunter, A.; Wright, P.; Cappelli, M.; Kasaboski, A.; Surh, L. Physician knowledge and attitudes towards molecular genetic (DNA) testing of their patients. *Clin. Genet.* **1998**, *53*, 447–455.
41. Haga, S.B.; Burke, W.; Ginsburg, G.S.; Mills, R.; Agans, R. Primary care physicians' knowledge of and experience with pharmacogenetic testing. *Clin. Genet.* **2012**, *82*, 388–394.
42. Edwards, Q.T.; Maradiegue, A.; Seibert, D.; Saunders-Goldson, S.; Humphreys, S. Breast cancer risk elements and nurse practitioners' knowledge, use, and perceived comfort level of breast cancer risk assessment. *J. Am. Acad. Nurse Pract.* **2009**, *21*, 270–277.
43. Bethea, J.; Qureshi, N.; Drury, N.; Guilbert, P. The impact of genetic outreach education and support to primary care on practitioner's confidence and competence in dealing with familial cancers. *Community Genet.* **2008**, *11*, 289–294.
44. Clyman, J.C.; Nazir, F.; Tarolli, S.; Black, E.; Lombardi, R.Q.; Higgins, J.J. The impact of a genetics education program on physicians' knowledge and genetic counseling referral patterns. *Med. Teach.* **2007**, *29*, e143–e150.
45. American Board of Medical Specialties. *ABMS Guide to Physician Specialties*; Elsevier: Maryland Heights, MO, USA, 2013; p. 44.

46. National Society of Genetic Counselors (NSGC). *Making Sense of Your Genes: A Guide to Genetic Counseling*; NSGC: Chicago, IL, USA, 2008; pp. 1–21.
47. American Board of Genetic Counseling Inc. about ABGC. Available online: http://www.abgc.net/About_ABGC/GeneticCounselors.asp/ (accessed on 6 February 2013).
48. National Society of Genetic Counselors (NSGC). *2012 Professional Status Survey: Executive Summary*; NSGC: Chicago, IL, USA, 2012; pp. 1–15.
49. McPherson, E.; Zaleski, C.; Benishek, K.; McCarty, C.A.; Giampietro, P.F.; Reynolds, K.; Rasmussen, K. Clinical genetics provider real-time workflow study. *Genet. Med.* **2008**, *10*, 699–706.
50. The Physician Workforce: Projections and Research into Current Issues Affecting Supply and Demand. Available online: http://bhpr.hrsa.gov/healthworkforce/reports/physwfissues.pdf (accessed on 15 February 2013).
51. Cooksey, J.A.; Forte, G.; Benkendorf, J.; Blitzer, M.G. The state of the medical geneticist workforce: Findings of the 2003 survey of American Board of Medical Genetics certified geneticists. *Genet. Med.* **2005**, *7*, 439–443.
52. Collins, F.S. Faith and the human genome. *Perspect. Sci. Christian Faith* **2003**, *55*, 142–153.
53. Belmont, J.; McGuire, A.L. The futility of genomic counseling: Essential role of electronic health records. *Genome Med.* **2009**, *1*, e48.
54. Ullman-Cullere, M.H.; Mathew, J.P. Emerging landscape of genomics in the electronic health record for personalized medicine. *Hum. Mutat.* **2011**, *32*, 512–516.
55. Ginsburg, G.S.; Willard, H.F. Genomic and personalized medicine: Foundations and applications. *Transl. Res.* **2009**, *154*, 277–287.
56. Overby, C.L.; Kohane, I.; Kannry, J.L.; Williams, M.S.; Starren, J.; Bottinger, E.; Gottesman, O.; Denny, J.C.; Weng, C.; Tarczy-Hornoch, P.; *et al.* Opportunities for genomic clinical decision support interventions. *Genet. Med.* **2013**, *15*, 817–823.
57. Kullo, I.J.; Jarvik, G.P.; Manolio, T.A.; Williams, M.S.; Roden, D.M. Leveraging the electronic health record to implement genomic medicine. *Genet. Med.* **2013**, *15*, 270–271.
58. Osheroff, J.A.; Teich, J.M.; Middleton, B.; Steen, E.B.; Wright, A.; Detmer, D.E. A roadmap for national action on clinical decision support. *J. Am. Med. Inform. Assoc.* **2007**, *14*, 141–145.
59. Wright, A.; Sittig, D.F.; Ash, J.S.; Feblowitz, J.; Meltzer, S.; McMullen, C.; Guappone, K.; Carpenter, J.; Richardson, J.; Simonaitis, L.; *et al.* Development and evaluation of a comprehensive clinical decision support taxonomy: Comparison of front-end tools in commercial and internally developed electronic health record systems. *J. Am. Med. Inform. Assoc.* **2011**, *18*, 232–242.
60. Bright, T.J.; Wong, A.; Dhurjati, R.; Bristow, E.; Bastian, L.; Coeytaux, R.R.; Samsa, G.; Hasselblad, V.; Williams, J.W.; Musty, M.D.; *et al.* Effect of clinical decision-support systems: A systematic review. *Ann. Intern. Med.* **2012**, *157*, 29–43.
61. Jaspers, M.W.; Smeulers, M.; Vermeulen, H.; Peute, L.W. Effects of clinical decision-support systems on practitioner performance and patient outcomes: A synthesis of high-quality systematic review findings. *J. Am. Med. Inform. Assoc.* **2011**, *18*, 327–334.
62. Chaudhry, B.; Wang, J.; Wu, S.; Maglione, M.; Mojica, W.; Roth, E.; Morton, S.C.; Shekelle, P.G. Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Ann. Intern. Med.* **2006**, *144*, 742–752.

63. Mitchell, J.A.; Gerdin, U.; Lindberg, D.; Lovis, C.; Martin-Sanchez, F.J.; Miller, R.A.; Shortliffe, E.H.; Leong, T.-Y. 50 Years of informatics research on decision support: What's next. *Methods Inf. Med.* **2011**, *50*, 525–535.
64. What is Meaningful Use? | Policy Researchers & Implementers | HealthIT.gov. Available online: http://www.healthit.gov/policy-researchers-implementers/meaningful-use (accessed on 8 November 2013).
65. Hunt, D.L.; Haynes, R.B.; Hanna, S.E.; Smith, K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: A systematic review. *JAMA* **1998**, *280*, 1339–1346.
66. Bates, D.; Kuperman, G. Ten commandments for effective clinical decision support: Making the practice of evidence-based medicine a reality. *J. Am. Med. Inform. Assoc.* **2003**, *10*, 523–530.
67. Kawamoto, K.; Houlihan, C.A.; Balas, E.A.; Lobach, D.F. Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. *Br. Med. J.* **2005**, *330*, e765.
68. Emery, J.; Morris, H.; Goodchild, R.; Fanshawe, T.; Prevost, A.T.; Bobrow, M.; Kinmonth, A.L. The GRAIDS trial: A cluster randomised controlled trial of computer decision support for the management of familial cancer risk in primary care. *Br. J. Cancer* **2007**, *97*, 486–493.
69. Collins, F.S. A Brief Primer on Genetic Testing. Available online: http://www.genome.gov/10506784/ (accessed on 8 February 2013).
70. NIH Office of Rare Diseases Research (ORDR). Undiagnosed Diseases Program. Available online: http://rarediseases.info.nih.gov/Resources.aspx?PageID=31/ (accessed on 20 February 2013).
71. Cirulli, E.T.; Goldstein, D.B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* **2010**, *11*, 415–425.
72. Bamshad, M.J.; Ng, S.B.; Bigham, A.W.; Tabor, H.K.; Emond, M.J.; Nickerson, D.A.; Shendure, J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **2011**, *12*, 745–755.
73. Berg, J.S.; Khoury, M.J.; Evans, J.P. Deploying whole genome sequencing in clinical practice and public health: Meeting the challenge one bin at a time. *Genet. Med.* **2011**, *13*, 499–504.
74. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD, USA), Online Mendelian Inheritance in Man, OMIM®. Available online: http://omim.org/ (accessed on 8 February 2013).
75. Genetics Home Reference Waardenburg syndrome. Available online: http://ghr.nlm.nih.gov/condition/waardenburg-syndrome/ (accessed on 20 February 2013).
76. Lalloo, F.; Evans, D.G. Familial breast cancer. *Clin. Genet.* **2012**, *82*, 105–114.
77. Drohan, B.; Roche, C.A.; Cusack, J.C.; Hughes, K.S. Hereditary breast and ovarian cancer and other hereditary syndromes: Using technology to identify carriers. *Ann. Surg. Oncol.* **2012**, *19*, 1732–1737.
78. Schwartz, M.D.; Valdimarsdottir, H.B.; DeMarco, T.A.; Peshkin, B.N.; Lawrence, W.; Rispoli, J.; Brown, K.; Isaacs, C.; O'Neill, S.; Shelby, R.; *et al.* Randomized trial of a decision aid for BRCA1/BRCA2 mutation carriers: Impact on measures of decision making and satisfaction. *Health Psychol.* **2009**, *28*, 11–19.

79. Glasspool, D.W.; Oettinger, A.; Smith-Spark, J.H.; Castillo, F.C.; Monaghan, V.E.L.; Fox, J. Supporting medical planning by mitigating cognitive load. *Methods Inf. Med.* **2007**, *46*, 636–640.
80. Reference, G.H. Cystic fibrosis. Available online: http://ghr.nlm.nih.gov/condition/cystic-fibrosis/ (accessed on 20 February 2013).
81. Srinivasan, B.S.; Evans, E.A.; Flannick, J.; Patterson, A.S.; Chang, C.C.; Pham, T.; Young, S.; Kaushal, A.; Lee, J.; Jacobson, J.L.; *et al.* A universal carrier test for the long tail of Mendelian disease. *Reprod. Biomed. Online* **2010**, *21*, 537–551.
82. Aithal, G.P.; Day, C.P.; Kesteven, P.J.L.; Daly, A.K. Association of polymorphisms in the cytochrome P450 CYP2C9 with warfarin dose requirement and risk of bleeding complications. *Early Rep.* **1999**, *353*, 717–719.
83. Secretary's Advisory Committee on Genetics, Health, and Society (SACGHS). *Realizing the Potential of Pharmacogenomics: Opportunities and Challenges*; SACGHS: Bethesda, MD, USA, 2008.
84. Gage, B.; Eby, C.; Johnson, J. Use of pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. *Clin. Pharmacol. Ther.* **2008**, *84*, 326–331.
85. Pulley, J.M.; Denny, J.C.; Peterson, J.F.; Bernard, G.R.; Vnencak-Jones, C.L.; Ramirez, A.H.; Delaney, J.T.; Bowton, E.; Brothers, K.; Johnson, K.; *et al.* Operational implementation of prospective genotyping for personalized medicine: The design of the Vanderbilt PREDICT project. *Clin. Pharmacol. Ther.* **2012**, *92*, 87–95.
86. Zeisel, S.H. Choline: Critical role during fetal development and dietary requirements in adults. *Annu. Rev. Nutr.* **2006**, *26*, 229–250.
87. Brody, L.C.; Conley, M.; Cox, C.; Kirke, P.N.; McKeever, M.P.; Mills, J.L.; Molloy, A.M.; O'Leary, V.B.; Parle-McDermott, A.; Scott, J.M.; *et al.* A polymorphism, R653Q, in the trifunctional enzyme methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase/formyltetrahydrofolate synthetase is a maternal genetic risk factor for neural tube defects: Report of the Birth Defects Res. *Am. J. Hum. Genet.* **2002**, *71*, 1207–1215.
88. Masys, D.R.; Jarvik, G.P.; Abernethy, N.F.; Anderson, N.R.; Papanicolaou, G.J.; Paltoo, D.N.; Hoffman, M.A.; Kohane, I.S.; Levy, H.P. Technical desiderata for the integration of genomic data into Electronic Health Records. *J. Biomed. Inform.* **2012**, *45*, 419–422.
89. Hoffman, M.A. The genome-enabled electronic medical record. *J. Biomed. Inform.* **2007**, *40*, 44–46.
90. Gottesman, O.; Kuivaniemi, H.; Tromp, G.; Faucett, W.A.; Li, R.; Manolio, T.A.; Sanderson, S.C.; Kannry, J.; Zinberg, R.; Basford, M.A.; *et al.* The Electronic Medical Records and Genomics (eMERGE) Network: Past, present, and future. *Genet. Med.* **2013**, *15*, 761–771.
91. McCarty, C.A.; Chisholm, R.L.; Chute, C.G.; Kullo, I.J.; Jarvik, G.P.; Larson, E.B.; Li, R.; Masys, D.R.; Ritchie, M.D.; Roden, D.M.; *et al.* The eMERGE Network: A consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med. Genomics* **2011**, *4*, e13.
92. National Center for Biotechnology Information ClinVar. Available online: http://www.ncbi.nlm.nih.gov/clinvar/ (accessed on 8 February 2013).
93. New NIH-Funded Resource Focuses on Use of Genomic Variants in Medical Care. Available online: http://www.nih.gov/news/health/sep2013/nhgri-25.htm (accessed on 8 October 2013).
94. Drohan, B.; Ozanne, E.M.M.; Hughes, K.S.S. Electronic health records and the management of women at high risk of hereditary breast and ovarian cancer. *Breast J.* **2009**, *15*, S46–S55.

95. Kawamoto, K.; Lobach, D. Proposal for fulfilling strategic objectives of the US roadmap for national action on decision support through a service-oriented architecture leveraging HL7 services. *J. Am. Med. Inform. Assoc.* **2007**, *14*, 146–155.
96. Standards & Interoperability (S&I) Framework—Health eDecisions Homepage. Available online: http://wiki.siframework.org/Health+eDecisions+Homepage/ (accessed on 8 November 2013).
97. Kawamoto, K.; Del Fiol, G.; Orton, C.; Lobach, D.F. System-agnostic clinical decision support services: Benefits and challenges for scalable decision support. *Open Med. Inform. J.* **2010**, *4*, 245–254.

CHAPTER 3

TECHNICAL DESIDERATA FOR THE INTEGRATION OF GENOMIC DATA WITH CLINICAL DECISION SUPPORT[1]

Introduction

Rapid genomic sequencing, including whole genome sequencing (WGS) and exome sequencing, is the future paradigm of clinical genetic testing [1]. With a patient's entire genome readily available to a clinician at the point of care, WGS may offer many benefits to traditional single gene testing. Currently, the effective use of single gene testing is inhibited by several factors including the need for clinical indication prior to ordering, the time delay between test ordering and the return of results, and financial constraints for single gene tests [2]. WGS will likely overcome many such barriers in the future. For instance, with WGS information readily available, pathogenic variants in disease-causing genes can be made known to the clinician much earlier in the decision-making process to aid in differential diagnosis [3]. Likewise, important but clinically under-utilized use cases such as pharmacogenomics may now become more clinically and financially feasible than

[1] A revised version of this chapter has been submitted to the *Journal of Biomedical Informatics* as a commentary. Co-authors of this submitted manuscript include Dr. Karen Eilbeck, Dr. Laurence Meyer, Dr. Guilherme Del Fiol, and Dr. Kensaku Kawamoto.

under the current genetic testing paradigm [4]. Indeed, the cost and value of WGS is now at a point at which health care providers are sequencing a patient's genome for unique clinical scenarios [5,6]. As a result, it may be soon when WGS becomes a larger part of routine clinical care [1].

Challenge of genome data in the clinical setting

While WGS offers many opportunities to enhance clinical care, were it to be made widely available for routine clinical care today, the effective use of WGS information would be hindered by significant barriers. These barriers include inadequate laboratory reporting methods, the complexity of genetic analysis, lack of physician proficiency in genetic analysis, and the insufficient number of genetics professionals in the workforce [7,8]. As clinicians are already burdened with significant time constraints, adding an additional layer of WGS information that they are required to review, integrate with other clinical parameters, and translate into appropriate clinical actions is unlikely to be successful without assistance [9]. This challenge is amplified by the rapidly evolving nature of our understanding of the genome and its clinical implications [10].

Potential of clinical decision support

Clinical decision support (CDS) integrated into the clinician's workflow provides a practical solution to allow clinicians to provide effective clinical care using genomic information [11,12]. CDS entails providing clinicians, patients, and other healthcare stakeholders with pertinent knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and healthcare [13]. Examples

of point-of-care CDS include medication dosing support, order facilitators, alerts and reminders, relevant information display, expert systems, and workflow support [14]. CDS knowledge, which supports these types of CDS, is used to process clinical data to provide patient-specific advice, recommendations, or information. Figure 3.1 provides summary of CDS types and potential examples of WGS-enabled CDS.

When developed and implemented properly, CDS has the ability to process large amounts of complex data, such as WGS data, and present actionable, evidence-based recommendations to clinicians at the point of care [15]. In doing so, CDS has been shown to be effective in reducing errors, improving clinician performance, and ultimately improving the quality of care in clinical settings [16]. Furthermore, CDS is able to translate research discoveries into clinical care much more efficiently than other traditional methods of knowledge translation [17]. Indeed, CDS may be essential to meeting the demands of WGS at the point of care [18].

Masys *et al.* desiderata for integration of genomic information with EHRs

To provide guidance on how to integrate genomic information within the EHR, Masys *et al.* [19] developed a set of guiding principles for the technical integration of genomic information into the electronic health record (EHR). Their summarized desiderata is outlined in Figure 3.2. While this paper provides a strong foundation for integration of genomic information within the EHR, in our experience developing CDS capabilities for WGS, we found that it did not fully address all the needs of WGS integration with CDS.

To address this need, we assembled a core group of domain experts in genomics

and CDS to define additional desirable functional characteristics for CDS capable of incorporating WGS at the point of care within an EHR. These additional requirements were then validated and assessed for importance among a larger group of domain experts in genomics and CDS. This effort is intended to augment the original Masys *et al.* desiderata and provide further guidance to system developers on important requirements to consider when developing health IT systems and CDS for WGS information. Of note, the requirements described in this desiderata represent current and potential needs according to our current understanding. Nevertheless, future research and development may require additional requirements to be added or compel current requirements to be removed from the desiderata. As such, this initial set of requirements should be viewed as an evolving set of requirements that informs and is informed by ongoing research and development in this field.

Methods

Development of additional desiderata

A core group of domain experts in genomics and CDS was assembled to review needs for WGS CDS. This group of domain experts derived a set of additional desiderata by consensus. The members of the core group consisted of the authors of this manuscript. Their credentials are described in Appendix A. The core group iteratively refined the proposed requirements with multiple rounds of revisions until an agreed upon initial set of desiderata was available for external review.

Input from community domain experts

We sought to improve, refine, and validate our initial set of requirements by seeking additional input and qualitative feedback from a wider community of domain experts in genomics and CDS. To obtain this feedback, an anonymous, internal review board (IRB) approved survey was distributed electronically to e-mail discussion lists of relevant expert groups. Participants included members of the HL7 Clinical Genomics Workgroup, HL7 CDS Workgroup, AMIA Genomics Workgroup, AMIA CDS Workgroup, Open Source Electronic Health Record Agent (OSEHRA) Genomics Workgroup, the developers of ClinVar, University of Utah Program in Personalized Health Care, and our own professional contacts. The survey instrument was available from 8/13/2013 through 8/30/2013.

Survey instrument

The survey data were collected and managed using the REDCap electronic data capture tool hosted at the University of Utah [20]. The survey consisted of a brief background followed by a demographic question assessing the type of expertise the survey participant held (genomics, CDS, or both). Each requirement was listed and summarized on a separate page with a 5-point Likert scale assessing the participant's opinion regarding the importance of each requirement (very important, important, neither important nor unimportant, unimportant, very unimportant, and unsure/no opinion/blank). Additionally, each requirement had a field to allow the participant to provide a comment. The survey concluded with a solicitation for general comments and the option for study participants to provide their email to be contacted again, if desired. The full survey is available in

Appendix B. For a participant's response to be included in the final analysis, the study participant had to consent to the survey and complete greater than 50% of the Likert-scale questions.

Data analysis and presentation

Survey responses were analyzed and visualized using spreadsheet software. To assess significance, Likert responses for each requirement were grouped into important (very important and important) and nonimportant (neither important nor unimportant, unimportant, very unimportant). Likert scale responses for each requirement are presented as graphs in this manuscript (see Figure 3.1). We used a chi square test with expected values based on an even selection of all values. Qualitative feedback was reviewed by the core panel, and feedback was incorporated into the desiderata where appropriate. Participant feedback, resulting modifications, and core group responses are available at http://tinyurl.com/nxjxr8v.

Results

A total of 108 surveys were started, of which 45 (42%) were excluded because less than 50% of Likert responses were completed. A total of 63 (58%) responses were included in the analysis; in these cases, almost all Likert questions were fully completed. Of the included participants, 27% were experts in genomics only, 49% were experts in CDS only, and 24% were experts in both CDS and genomics. The majority of included surveys (79%) were completed within the first week of the survey. All proposed requirements were judged to be important (which includes important and very important) by the community of experts. All results were found to be highly significant ($p<10\text{-}9$). The final desiderata is

summarized in Figure 3.3, listed in order of importance as assessed by the community of experts. The original order of requirements presented in the survey to study participants are available in Appendix B. As this is an effort to expand, not replace, the original Masys *et al.* requirements, the desiderata numbering will continue with the number eight and follow the same summary and description format for each requirement. Tables summarizing in graphical format the participant responses for each requirement are available at http://tinyurl.com/nxjxr8v.

CDS knowledge must have the potential to incorporate multiple genes and clinical information (Requirement #8)

A relatively small number of Mendelian diseases, such as cystic fibrosis and sickle-cell anemia, are affected by variants within a single gene responsible for producing the characteristic phenotype. As a result, such cases are fairly straightforward to assess. With nearly every human condition affected one way or another by a genetic influence, most diseases, in particular common diseases, are caused or affected by multiple genetic influences and environmental factors. For example, there are potentially hundreds of genetic loci contributing to type 2 diabetes risk [21]. In order to provide an accurate risk assessment and decision support, all relevant genetic loci need to be considered, as well as any relevant clinical factors (e.g., age, weight, health history, and comorbidities) and environmental influences (e.g., diet, physical activity, stress). Furthermore, it may not always be the case that all necessary data are in one central location. Therefore, CDS for the WGS must have the capacity to leverage and incorporate several pieces of information from multiple genomic and nongenomic data sources.

Keep CDS knowledge separate from the variant classification (Requirement #9)

Masys *et al.* describe the importance of separating molecular observations (e.g., DNA sequence) from variant classification (e.g., pathogenicity classifications) due to the need to update variant interpretation as genome knowledge changes and grows over time. To illustrate, one study found that over a seven year period, 14.5% of reported variant classifications had to be reclassified [22]. Likewise, it is also essential to separate CDS knowledge from both molecular observations and variant classification. CDS must have the ability to manage evolving and frequently changing gene variant interpretations efficiently without requiring changes to the underlying CDS knowledge each time a variant's classification changes. Separation of CDS knowledge from variant interpretation allows CDS knowledge to be more efficiently handled and maintained.

Have the capacity to support multiple EHR platforms with various data representations with minimal modification (Requirement #10)

The reality of the health information environment in the US today is that multiple healthcare organizations use multiple EHR and health information management systems [23]. Often, these health information management solutions store and represent the same health information differently. This can be a challenge when trying to harness the information within different health IT systems in different organizations to provide CDS. Due to the need to distribute and share WGS enabled CDS knowledge across multiple organizations (see next requirement), the CDS architecture would ideally be EHR agnostic,

where CDS knowledge can be developed once, and then run consistently anywhere. A number of initiatives aimed at supporting this type of architecture are underway, including the Health eDecisions initiative, OpenCDS, the SMART platform, and the CDS Consortium, to name a few.

Support a large number of gene variants while simplifying the CDS knowledge to the extent possible (Requirement #11)

There are roughly 1200 known variants in the adenomatous polyposis coli (APC) gene, a gene associated with a rare form of colon cancer [24]. Likewise, there are nearly 2000 known variants in the cystic fibrosis gene CFTR [25]. Given the potentially high number of variants per gene, it may be inefficient to create CDS knowledge for every known variant in each disease-causing gene. Furthermore, as novel variants are discovered, it will be difficult to update CDS knowledge for every gene variant that is discovered. Therefore, to manage this complexity, variants with the same or similar clinical impact should be classified accordingly. CDS knowledge can then be simplified by developing logic or rules which leverages the variant interpretation rather than the specific variant. Nevertheless, in cases where a particular variant has a unique and clinically important impact or where machine learning CDS models could utilize individual variants or combinations of variants within a single gene, genetic information at the variant level should still be accessible to CDS knowledge. In summary, CDS knowledge can be greatly simplified by classifying variants into groups of common clinical impact, while still supporting inferencing at the individual variant level where necessary.

Leverage current and developing CDS and genomics infrastructure and standards (Requirement #12)

Both the CDS and genomics fields have benefited from extensive research and development over the years. Indeed, both fields have well developed infrastructure and standards to support its uses. Therefore, it is important to leverage these standards and infrastructure. Examples include using Human Gene Variation Society (HGVS) and dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) to represent specific molecular observations; American College of Medical Genetics and Genomics (ACMG) recommendations for variant classification; HL7 Clinical Genomics standards for the representation of genetic information; Arden Syntax, GELLO, GLIF3, GEM, or the HL7 CDS Knowledge Artifact Implementation Guide for CDS knowledge representation; the HL7 Decision Support Service standard and HL7 Infobutton standard for delivering CDS as a service; and open-source, standards-based resources such as OpenCDS. While many of these standards and resources may not be completely sufficient for meeting the needs of CDS for WGS, it still represents significant relevant effort. It will be important to leverage current and developing CDS and genomics infrastructure, standards, and knowledge.

Support a CDS knowledge base deployed at and developed by multiple independent organizations (Requirement #13)

With the potential for genomic information to impact nearly every clinical decision and the clinical application of genomics rapidly evolving, the time and cost for a single entity or organization to manually create and update CDS knowledge will be prohibitive. Indeed, no one organization will be able to author and manage all CDS knowledge for all

WGS use cases. Furthermore, there must be an efficient and scalable mechanism to consistently modify CDS knowledge everywhere it is deployed. Ideally, a standardized CDS infrastructure would allow multiple health care organizations, public or private entities, or individuals to create, publish, and distribute CDS knowledge efficiently to multiple consuming health care organizations. Such an approach will allow a specialized entity (e.g., pharmacogenomics experts) to develop and manage CDS knowledge and subsequently distribute to 'subscribing' health care organizations. With an ecosystem of CDS knowledge developed independently by multiple content developers, it becomes more feasible for health care organizations to have affordable access to comprehensive, up-to-date, and accurate CDS knowledge for the entire genome.

Access and transmit only the genomic information necessary for CDS (Requirement #14)

The separation of CDS knowledge from molecular observations and variant interpretations will require relevant genetic information being accessed and sent to a CDS engine (or equivalent) for processing. It will be inefficient and insecure to transmit an entire genome file for every CDS knowledge. The processing capacity required to transmit and sift through an entire genome for CDS knowledge will hinder the ability to provide CDS at the point of care in real-time. Furthermore, HIPAA requires that only the minimum protected health information needed to satisfy a particular purpose or carry out a function be used or transmitted [26]. Therefore, a CDS architecture must only transmit the relevant genes and any associated molecular observations, variant interpretations, and associated clinical information.

Summary of results

Figure 3.4 displays the proportion of each survey response for all requirements in the desiderata.

Discussion

Summary of findings

To guide development of CDS for WGS, we identified several additional requirements that were not addressed in a previously published work on integrating genomic information with EHRs [19]. While the Masys *et al.* desiderata primarily focused on the integration of genomic information within the EHR, this desiderata focuses largely on the integration of genomic information with CDS capabilities. The combination of both desiderata is important to leverage WGS within EHRs using CDS.

Insights from the community of domain experts

In general, the community of domain experts provided encouraging and insightful comments. The majority of comments for each requirement were largely supportive of the proposed requirements, often reiterating the stated importance and occasionally offering suggestions on how to accomplish the requirement. Occasionally a participant would express doubt that the proposed requirement could actually be accomplished in the current health IT environment, even if they had selected the requirement as important. These reasons for doubt varied by requirement and expertise. Where appropriate, comments were provided to justify one's opposition to a proposed requirement. In all, the comments were very useful in understanding the thoughts and intensions of a participant's particular

responses to the Likert questions.

While the community of domain experts generally agreed that the proposed desiderata were all important, we found some of the feedback interesting. In particular, we were surprised by the relative low ranking in importance of desiderata #14 ('Access and transmit only the genomic information necessary for CDS') compared to the other requirements. As this requirement addresses issues of privacy and security, a sensitive topic in health care, we did not expect respondents to rank this the least important requirement relative to the other requirements. However, in reviewing the qualitative feedback on this requirement, we discovered several respondents to be in disagreement that the HIPAA 'Minimum Necessary' requirement would need to be applied in this case. Nevertheless, upon consulting with our institution's health information security experts regarding this scenario, they reaffirmed this proposed need to access and transmit only the necessary genetic information needed for CDS. Furthermore, while this requirement may be important given current technical capabilities, it may be possible that such concerns become less important in the future as relevant technologies and technology and laws evolve. This possibility could also be reflected in the comparatively lower importance rating given to this requirement by the study participants.

Assumptions and implications

In developing the desiderata we focused on CDS delivered based on well-defined knowledge. In other words, we assumed that gene variants and their clinical implications were already established. However, there is also a clinical need for genome analysis tools such as Omicia Opal, SVBio, and others to support the manual identification of uncommon

or novel genetic variants suspected of causing rare phenotypes or orphan diseases [27]. In this analysis, we did not focus on this type of genome analysis tool, which could be considered a CDS tool, because they will likely not be integrated with routine clinical workflow as automatic CDS.

Another assumption made in developing the desiderata is that the variant classification is accurate and consistent between genome variant knowledge bases. In reality, variant knowledge bases are known to provide conflicting information, which can be problematic, particularly if a genome database is checking several genome variant knowledge bases for interpretation. However, as the genomics industry grows and matures over time, we anticipate that the quality and collaboration between these variant knowledge bases will improve. It is also possible that a centralized variant knowledge base, such as ClinVar, will manage all variant classifications [28]. Furthermore, it is important to consider the possibility that the same variant shared between two individuals could have different variant classifications based upon extrinsic factors. The implication of this complexity in variant classification is that the additional decision logic may need to be considered when assigning variant classifications.

Finally, we assumed that WGS data quality is consistent with other methods of genetic testing, currently this is not the case [29]. Often if a pathogenic variant is identified with a WGS test, the variant is confirmed using more reliable genetic testing such as Sanger sequencing [30]. Improvements in data quality will be necessary before WGS testing can be used as a stand-alone genome assessment method in the clinical setting. However, we also anticipate that current and future developments of sequencing technology and bioinformatics capabilities will improve genome data quality to a level acceptable for

clinical application.

Desired capabilities and correspondence to desiderata

There is much debate regarding the laboratory's role in recontacting a treating clinician each time a variant classification changes [31,32]. For some, the ability to alert treating clinicians to changes in variant classification, particularly from an unknown or benign classification to a pathogenic one, is an important responsibility. Indeed, this capability is the premise behind the GeneInsight software [22,33]. With regards to this study, alerting to changes to variant classification is built into the CDS functions represented by desiderata #9 and #11. Every time CDS is triggered, it can re-assess the variant classification and utilize the most recent classification available that time. One would not necessarily need to alert a clinician when a variant classification changes; rather, CDS knowledge could periodically check the patient's genome and provide the new interpretation and recommendations within the workflow, as necessary.

Furthermore, presentation of CDS results and recommendations to clinicians in an understandable and actionable way is another desirable capability. There is much research and efforts already underway in this space. The incorporation of best practices identified in this manner can be considered an additional aspect of desiderata #12 ('Leverage current and developing CDS and genomics infrastructure and standards').

Strengths and limitations

A major strength of this study is the development of the proposed desiderata by a diverse core panel of domain experts and the subsequent validation among a larger group

of domain experts. In particular, we invited groups with participants who (1) have experience developing and deploying CDS systems, and/or (2) are involved with the clinical application or interpretation of genomic information in health care. Another strength of our approach is the relative ranking of the desiderata, which provide some guidance on the relative importance of supporting these desired features when developing CDS capabilities for WGS information.

A shortcoming of the study is the use of a pragmatic research design. Specifically, this study only assessed community input once, as compared to multiple rounds of feedback in methods such as the Delphi method [34]. However, as this survey was anonymous and contact information was not provided by most participants, it was not feasible to have multiple rounds of input from the community domain experts. Nevertheless, the core panel did conduct iterative rounds of revisions based on the community experts' input. A second limitation of this study is that we only asked for the importance of desiderata, whereas we could have asked for additional information, such as feasibility and current availability. However, we felt that perceived need was of greatest importance. Also, we wished to maximize broad participation by keeping the survey relatively simple and short. Therefore, we believe the utilized study design maximized the amount of useful information obtained from the community of domain experts.

Another potential shortcoming is that we did not explicitly solicit participation from additional professional networks with expertise in this domain, such as the Electronic Medical Records and Genomics (eMERGE) Network and Pharmacogenomics Research Network (PGRN). However, we believe that the personal and professional networks used in this study constitute several individuals belonging to eMERGE, PGRN, and other

groups. Additionally, we received a large enough representative sample size from the participating networks and qualified individuals, such that additional participants would likely not have changed the results significantly.

Future direction

Given the important need to provide CDS capable of supporting WGS information at the point of care, and the requirements to support both the Masys *et al.* and these desiderata, a novel CDS architecture capable of supporting these requirements will need to be designed, developed, and evaluated. Due to the scope and complexity of integrating WGS information with CDS as outlined in the desiderata, we propose a CDS architecture that utilizes principles of a service-oriented architecture (SOA). SOA is a software design methodology based on a collection of separate, independent software components known as services, which are self-contained and have well-defined capabilities [35]. Accordingly, we are currently developing a prototype CDS architecture that is based on the principles of SOA and is capable of supporting WGS information in a manner consistent with the desiderata described in this manuscript and the Masys *et al.* paper. Once such a prototype is developed and tested, clinical scenarios will test the feasibility of this approach. Finally, once testing is complete, implementing as a pilot study will be important to validate the solution in a real clinical setting. As a result of these efforts, it is likely that additional requirements may be identified for inclusion in the desiderata. As such, this proposed set of desiderata should not be considered a final authoritative set, but rather a foundation upon which further requirements may be added in the future.

Conclusion

CDS to support WGS information at the point of care will likely be necessary to meet the clinical demands of the genome. However, the complexity of WGS information will require an approach for implementing CDS that is flexible and robust. We have added and validated several additional requirements to the Masys *et al.* desiderata which specifically focus on CDS needs for WGS information. We speculate that these additional desiderata will benefit the design and development of CDS approaches for supporting the full integration of WGS information into clinical care.

References

1. Bonetta L. Whole-genome sequencing breaks the cost barrier. *Cell*. 2010:917–919.

2. Weldon CB, Trosman JR, Gradishar WJ, Benson AB, Schink JC. Barriers to the use of personalized medicine in breast cancer. *J Oncol Pract*. 2012;8(4):e24–31. doi:10.1200/JOP.2011.000448.

3. Feero WG, Guttmacher AE, Collins FS. Genomic medicine - An updated primer. *N Engl J Med*. 2010;362(21):2001–2011.

4. McWilliam A, Lutter R, Nardinelli C. Health care savings from personalizing medicine using genetic testing: The case of warfarin. *AEI-Brookings Jt Cent Regul Stud*. 2006;(November):1–17.

5. Saunders CJ, Miller NA, Soden SE, *et al.* Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci Transl Med*. 2012;4(154):154ra135. doi:10.1126/scitranslmed.3004041.

6. Roach JC, Glusman G, Smit AFA, *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science*. 2010;328(5978):636–9. doi:10.1126/science.1186802.

7. Green RC, Rehm HL, Kohane IS. Clinical Genome Sequencing. In: *Genomic and Personalized Medicine, 2nd edition*.; 2013:102–122. doi:10.1016/B978-0-12-382227-7.00009-4.

8. Schrijver I, Aziz N, Farkas DH, *et al.* Opportunities and challenges associated with clinical diagnostic genome sequencing: A report of the association for molecular pathology. *J Mol Diagnostics*. 2012;14(6):525–540. doi:10.1016/j.jmoldx.2012.04.006.

9. Masys DR. Effects of current and future information technologies on the health care workforce. *Health Aff*. 2002;21(5):33–41.

10. Shirts BH, Parker LS. Changing interpretations, stable genes: Responsibilities of patients, professionals, and policy makers in the clinical interpretation of complex genetic information. *Genet Med*. 2008;10(11):778–83. doi:10.1097/GIM.0b013e31818bb38f.

11. Welch BM, Kawamoto K. Clinical decision support for genetically guided personalized medicine: A systematic review. *J Am Med Inform Assoc*. 2012;20(2):388–400. doi:10.1136/amiajnl-2012-000892.

12. Kawamoto K, Lobach DF, Willard HF, Ginsburg GS. A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. *BMC Med Inform Decis Mak*. 2009;9:17. doi:10.1186/1472-6947-9-17.

13. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. *J Am Med Informatics Assoc JAMIA*. 2007;14(2):141–145. doi:10.1197/jamia.M2334.Introduction.

14. Wright A, Sittig DF, Ash JS, *et al.* Development and evaluation of a comprehensive clinical decision support taxonomy: Comparison of front-end tools in commercial and internally developed electronic health record systems. *J Am Med Inform Assoc*. 2011;18(3):232–42. doi:10.1136/amiajnl-2011-000113.

15. Bennett C, Doub T, Selove R. EHRs connect research and practice: Where predictive modeling, artificial intelligence, and clinical decision support intersect. *Heal Policy Technol*. 2012:1–24. Available at: http://arxiv.org/abs/1204.4927. Accessed February 15, 2013.

16. Jaspers MWM, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: A synthesis of high-quality systematic review findings. *J Am Med Inform Assoc*. 2011;18(3):327–34. doi:10.1136/amiajnl-2011-000094.

17. Sarkar IN. Biomedical informatics and translational medicine. *J Transl Med*. 2010;8:22. doi:10.1186/1479-5876-8-22.

18. Ullman-Cullere MH, Mathew JP. Emerging landscape of genomics in the electronic health record for personalized medicine. *Hum Mutat*. 2011;32(5):512–

516. doi:10.1002/humu.21456.

19. Masys DR, Jarvik GP, Abernethy NF, *et al.* Technical desiderata for the integration of genomic data into Electronic Health Records. *J Biomed Inform*. 2012;45(3):419–22. doi:10.1016/j.jbi.2011.12.005.

20. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–81. doi:10.1016/j.jbi.2008.08.010.

21. Imamura M, Maeda S. Genetics of type 2 diabetes: The GWAS era and future perspectives. *Endocr J*. 2011;58(9):723–39.

22. Aronson SJ, Clark EH, Varugheese M, Baxter S, Babb LJ, Rehm HL. Communicating new knowledge on previously reported genetic variants. *Genet Med*. 2012;14(8):713–719. doi:10.1038/gim.2012.19.

23. Zhang NJ, Seblega B, Wan T, Unruh L, Agiro A, Miao L. Health information technology adoption in U.S. acute care hospitals. *J Med Syst*. 2013;37(2):9907. doi:10.1007/s10916-012-9907-2.

24. APC homepage - Colon cancer gene variant databases - Leiden Open Variation Database. Available at: http://chromium.liacs.nl/LOVD2/colon_cancer/home.php?select_db=APC. Accessed February 19, 2013.

25. Cystic Fibrosis Mutation Database: Statistics. Available at: http://www.genet.sickkids.on.ca/StatisticsPage.html. Accessed February 19, 2013.

26. US Department of Health and Human Services Office of the Secretary. *Standards for Privacy of Individually Identifiable Health Information; Final Rule*; 2000:45 CFR 164.502(b), 164.514(d).

27. Gullapalli RR, Lyons-Weiler M, Petrosko P, Dhir R, Becich MJ, Laframboise W a. Clinical integration of next-generation sequencing technology. *Clin Lab Med*. 2012;32(4):585–99. doi:10.1016/j.cll.2012.07.005.

28. Riggs ER, Wain KE, Riethmaier D, *et al.* Towards a universal clinical genomics database: The 2012 International Standards for Cytogenomic Arrays Consortium Meeting. *Hum Mutat*. 2013;34(6):915–9. doi:10.1002/humu.22306.

29. Lam HYK, Clark MJ, Chen R, *et al.* Performance comparison of whole-genome sequencing platforms. *Nat Biotechnol*. 2012;30(1):78–82. doi:10.1038/nbt.2065.

30. Chen Y, Zhang Q, Shen T, *et al.* Comprehensive mutation analysis by whole-

exome sequencing in 41 Chinese families with Leber congenital amaurosis. *Invest Ophthalmol Vis Sci*. 2013;54(6):4351–7. doi:10.1167/iovs.13-11606.

31. Hirschhorn K, Fleisher L. Duty to re-contact. *Genet Med.* 1999;1(4):171–172.

32. Richards CS, Bale S, Bellissimo DB, *et al.* ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med*. 2008;10(4):294–300. doi:10.1097/GIM.0b013e31816b5cae.

33. Aronson S, Clark E, Babb L. The GeneInsight Suite: A platform to support laboratory and provider use of DNA based genetic testing. *Hum Mutat*. 2011;32(5):532–536. doi:10.1002/humu.21470.

34. Linstone HA, Turoff M. *The Delphi Method- Techniques and Applications*. Addison-Wesley Pub. Co.; 2002.

35. Juneja G, Dournaee B, Natoli J, Birkel S. Improving performance of healthcare systems with service oriented architecture. *InfoQ*. 2008:1–15. Available at: http://www.infoq.com/articles/soa-healthcare.

CDS type	Clinical genomics example
Medication dosing support	CDS automatically adjusts warfarin dosing as a result of known alleles in the VKORC1 and CYP2C9 genes.
Order facilitators	An order for colonoscopy is recommended at a younger age as a result of known pathogenic mutations in genes associated with colon cancer.
Alerts and reminders	During medication ordering, gene variants known to affect drug pharmacokinetics are checked and clinicians are alerted to potential gene-drug interactions.
Relevant information display	Context aware infobuttons in the problem list leverage genome data to provide genetic risk information for a patient with breast cancer.
Expert systems	The EHR provides a 10-year cardiovascular disease risk score based on clinical, environmental, and genetic risk factors.
Workflow support	The EHR schedules a genetic counseling consultation during prenatal visit due to presence of an X-linked disease gene variant.

Figure 3.1: Potential examples of CDS leveraging WGS data

1. Maintain separation of primary molecular observations from the clinical interpretations of those data
2. Support lossless data compression from primary molecular observations to clinically manageable subsets
3. Maintain linkage of molecular observations to the laboratory methods used to generate them
4. Support compact representation of clinically actionable subsets for optimal performance
5. Simultaneously support human-viewable formats and machine-readable formats in order to facilitate implementation of decision support rules
6. Anticipate fundamental changes in the understanding of human molecular variation
7. Support both individual clinical care and discovery science

Figure 3.2: Desiderata for the integration of genomic data into EHRs described by Masys *et al.*

8. CDS knowledge must have the potential to incorporate multiple genes and clinical information.
9. Keep CDS knowledge separate from variant classification.
10. CDS knowledge must have the capacity to support multiple EHR platforms with various data representations with minimal modification.
11. Support a large number of gene variants while simplifying the CDS knowledge to the extent possible.
12. Leverage current and developing CDS and genomics infrastructure and standards.
13. Support a CDS knowledge base deployed at and developed by multiple independent organizations.
14. Access and transmit only the genomic information necessary for CDS.

Figure 3.3: Additional desiderata for the technical integration of WGS with CDS

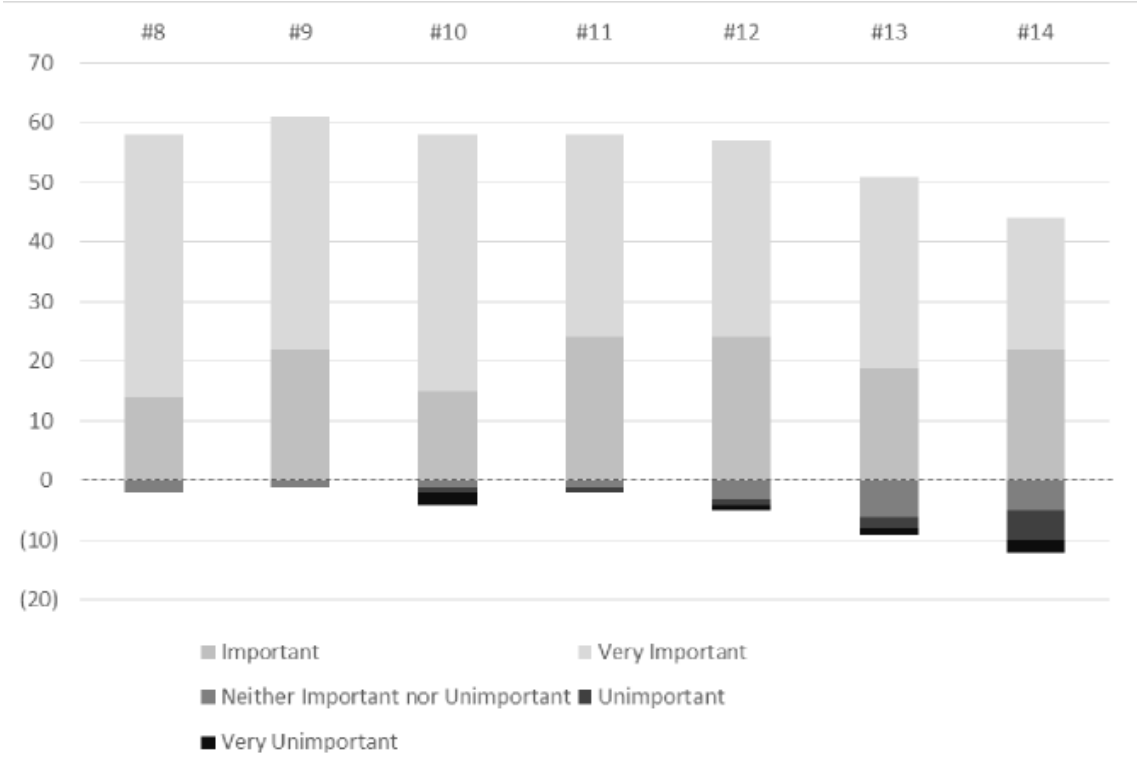


Figure 3.4: The proportion of each survey response for all requirements in the desiderata

CHAPTER 4

A PROPOSED CLINICAL DECISION SUPPORT ARCHITECTURE CAPABLE OF SUPPORTING WHOLE GENOME SEQUENCE INFORMATION[2]

Introduction

The use of whole genome sequence (WGS) information for routine clinical care will greatly enable the possibilities of personalized medicine, which include (1) improving diagnostic accuracy and disease characterization, (2) targeting therapies to individuals, (3) identifying and preventing disease among high-risk individuals, (4) improving healthcare efficiency, and (5) reducing unnecessary costs [1,2]. With genomic information readily available to clinicians at the point of care, many of these goals can be realized. Indeed, significant investment has been made to improve genome sequencing technology and reduce sequencing costs, making it easier to obtain a patient's WGS for clinical care [3]. As a result, WGS information is now being used in the clinical setting for rare, undiagnosed disorders [4–7]. If current trends continue, it is anticipated that WGS information will soon be available for routine clinical care, thus enabling personalized medicine on a widespread

[2] Reprinted with permission of Welch BM, Loya SR, Eilbeck K, Kawamoto K. A Proposed Clinical Decision Support Architecture Capable of Supporting Whole Genome Sequence Information. *J. Pers. Med.* 2014; 4(2):176-199.

scale [8].

While this is an intriguing prospect for patients, clinicians, and researchers, significant barriers exist which may hinder the effective use of WGS information in a routine clinical care setting. These barriers include (1) static laboratory reports intended for human consumption, (2) the complexity of genetic analysis, (3) limited physician proficiency in genetics, and (4) the lack of genetics professionals in the clinical workforce [9]. These barriers, if not overcome, will likely hinder the ability of clinicians to provide personalized medicine using WGS information. Although there may be several approaches to overcome these barriers, we believe clinical decision support (CDS) provided within the clinical workflow provides the greatest opportunity to enable effective use of WGS information in a routine clinical setting [9,10].

CDS entails providing clinicians, patients, and other healthcare stakeholders with pertinent knowledge and/or person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care [11]. Examples of CDS include medication dosing support, order facilitators, point of care alerts and reminders, relevant information display, expert systems, and workflow support [12]. Research on CDS has been conducted for several decades with established literature defining features that contribute to successful CDS interventions [13,14]. To be effective, it is essential that CDS for WGS information follow these proven CDS practices and approaches; in particular, the integration of CDS with the clinician's electronic health record (EHR) [9].

State of the art

While CDS research is a well-established field, research on CDS for genetically-guided personalized medicine is a much younger, but growing, field. In a systematic review of CDS interventions for genetically-guided personalized medicine, Welch and Kawamoto identified 16 primary research articles describing CDS interventions using genetic information between 1990 to 2011 [15]. The majority of these CDS interventions tended to be stand-alone applications, which required re-entry of a patient's clinical and genomic data by a clinician. Furthermore, these applications were largely limited to a single, or limited number, of genes (e.g., BRCA) [16]. Recently, Tarczy-Hornoch *et al.* conducted a review of clinical reporting approaches for WGS (and whole exome) information in the EHR, which are currently implemented at six healthcare organizations [17]. These healthcare organizations developed, implemented, and managed various approaches to EHR integration and CDS. However, the majority of these approaches were limited to static PDF reports (similar to pathology reports), and only two organizations leveraged active CDS capabilities of the local EHR. The authors acknowledge that active CDS will be necessary for WGS information and that more sophisticated informatics tools will be necessary to scale up to meet the challenges of WGS information [17]. A more detailed description of these CDS examples and how they compare to the work described in this manuscript can be found in the Discussion section. In general, literature on CDS for WGS information is still in its infancy [18].

Technical desiderata

Given the critical role health IT will play in overcoming the barriers of WGS information and the specific challenges inherent in using genomic information, Masys *et al.* developed a technical desiderata for the integration of genomic information with an EHR [19]. These requirements, which were developed by a panel of experts, illustrate important considerations that should be addressed when developing health IT applications capable of supporting genomic information (see Figure 4.1). Indeed, these desiderata are intended to overcome many of the barriers and challenges (also described in the Masys *et al.* manuscript) of using genomic information for clinical care.

While the Masys desiderata provide a strong framework for integrating genomic data with the EHR, additional requirements are desirable for the integration of genomic information with CDS. Indeed, we believe it will be essential that genomic data are not only available within the EHR, but provided in a way that is useful to clinicians through CDS [9]. To address this need, Welch *et al.* developed an additional desiderata, to augment the Masys *et al.* desiderata, specifically focused on the integration of genomic information with CDS (see Figure 4.2) [20]. This work also describes the barriers and challenges that these additional requirements attempt to address.

These additional desiderata, when used together with the Masys *et al.* desiderata, can provide a foundation to guide research and development on CDS for WGS information. As there are many barriers inherent in leveraging WGS information for CDS [9], incorporating these desiderata into the design and development process may help system developers overcome the challenges of using WGS information.

Study objective

Given the importance that CDS will play in realizing personalized medicine through WGS information, and the early stage of research and development in this domain [15,17,18], we put forth a theoretical CDS architecture based upon the technical desiderata and approaches utilized in prior work [15,17]. Indeed, this manuscript lays out the conceptual design of a proposed architecture and describes how each component of the architecture attempts to meet the requirements described in the technical desiderata. It is our intent to put forward this proposed architecture as a foundational reference for research and development on CDS for WGS information in the future.

Methods

We have leveraged our collective experience in the domains of genetics, bioinformatics, and clinical informatics to propose a CDS architecture capable of supporting WGS information at the point of care. This manuscript, while describing the need for a particular approach or components, does not attempt to define the architecture components in sufficient detail necessary for implementation. Rather, this manuscript provides a business case and justification for the approaches and components used in this architecture.

Architecture overview

Given the complexity of WGS information, the success of CDS in the genomic age will likely require an architecture that separates key capabilities into independently managed component parts [21]. As such, we advocate the use of a service-oriented

architecture (SOA) as a design principle for our proposed CDS architecture. SOA is a software design methodology based on the interaction of separate, independent software components, known as services [22]. A service is a self-contained component which has well-defined, understood capabilities. SOA supports the reusability and standardization of processes, allowing for independent evolution and modifications to a particular service, reducing the burden of change on the overall system [23]. Because of the vast number of disparate health IT systems, the application of SOA principles offers several benefits to health care [24]. Indeed, research and development on SOA for CDS has led to several health IT standards and applications [25–28]. Furthermore, SOA-based CDS is currently under consideration for EHR certification criteria related to Stage III Meaningful Use guidelines [29].

SOA CDS for WGS information

While SOA offers many benefits to health IT and CDS, we believe it will be necessary for WGS-enabled CDS [21]. Indeed, SOA can provide the agility needed to keep up with the rapidly evolving genomics knowledge base [30]. Furthermore, SOA allows for the scalability that is needed to handle the breadth of genomic applications in health care, particularly across multiple independent health care organizations [31]. In contrast, were a health care organization to develop and maintain their own CDS knowledge for WGS information, they could become overwhelmed by the time and cost of creating, managing, and updating the CDS knowledge base for the entire genome [32]. This would be particularly challenging for the majority of health care organizations that have a limited clinical genomics presence [33]. Indeed, we believe it would be prudent to separate key

components into independently managed services, which can be optimally maintained by third-party organizations.

A SOA-based CDS architecture for the WGS information is an extension of previous efforts on SOA-based CDS in general [28] and early examples of CDS for genetic information [34–36]. The services and components required in our proposed architecture consist of genome sequencing and annotation, genome databases, genome variant knowledge bases, CDS knowledge base, CDS controller, and the EHR (see Figure 4.3). A glossary of terms and brief descriptions is available in Appendix C. While some of these services and components are already available, some will need to be developed or enhanced to support a SOA-based approach. In subsequent sections of this manuscript we describe each component in further detail, how they interact with each other, and enhancements that may be necessary.

Genome sequencing and annotation pipeline

The first step in the entire process is to obtain the patient's genome sequence, for either the whole genome, the exome, a gene panel or a more targeted, smaller subset of the genome. For a whole genome sequence, when compared to a reference genome, there are roughly three million single nucleotide variants per comparison. Two file formats for representing a patient's set of genome variants include the variant call format (VCF) and genome variant format (GVF) [37,38]. Both formats are able to represent various sequence and structural variations in the genome, such as single nucleotide polymorphisms, indels, and substitutions.

Genome annotation

Once the variants in the genome have been identified, it is necessary to prioritize variants that may have relevant phenotypic impacts. There are several sequential steps to variant annotation, which is referred to as the 'annotation pipeline.' Initially, this process identifies variants occurring within known or predicted genes, regulatory regions, protein coding sequences, or splice sites. Variants which occur within genes are assessed for clinical impact using curated genome variant knowledge bases, such as the Human Genome Mutation Database (HGMD), Online Mendelian Inheritance in Man (OMIM), ClinVar, and other locus-specific mutation databases [39–42]. Additionally, computational interpretation approaches such as Variant Annotation, Analysis & Search Tool (VAAST), SIFT (http://sift.jcvi.org/) and Polymorphism Phenotyping (PolyPhen) can be employed to predict variant pathogenicity based upon the impact on the gene's translational product [43–46]. Finally, gene functions, links to external knowledge resources, and other variant metadata can also be included. The annotation pipeline can be developed internally by the organization sequencing the patient's genome or using a service provided by a private company specializing in genome annotation services [47–49]. Currently, the entire sequencing and annotation pipeline is typically managed by a pathology laboratory. However, as genome sequencing technology advances, some speculate that this process could occur in the clinic [50]. In such cases, the proposed CDS architecture could still support this approach as long as this component interacts, in a similar way, with the other components of the architecture, namely the genome variant knowledge bases and genome database.

Genome variant knowledge base

A key part of the genome annotation process is to identify genome variants and assign a clinical impact, if known. A genome variant knowledge base is a repository of known genome variants and associated clinical interpretations of that variant. During the annotation pipeline, genome variant knowledge bases are ascertained for pre-existing knowledge on variants. There are many types of genome variant knowledge bases, which include (1) privately-controlled knowledge bases, such as the Human Gene Mutation Database (HGMD) [39]; (2) open access, locus-specific knowledge bases, such as those created using the Leiden Open Variation Database (LOVD) [42]; (3) proprietary knowledge bases, typically owned and managed by genetic testing laboratories, who maintain exclusive access [51]; and (4) publicly available, centrally-managed repositories, such as ClinVar [41]. Typically, when a new variant is discovered, or new information about a known variant is made available, this information will be recorded in one or more of these knowledge bases. Furthermore, curators may monitor publications and reports in order to update a knowledge base accordingly.

ClinVar, which is a publicly available central resource managed by the National Library of Medicine, represents a model wherein genome knowledge bases and laboratories (described above) can upload their expertly curated knowledge into one location. Previously, genome annotators may have had to use several different genome variant knowledge bases and pay to access particular knowledge. Furthermore, with a participatory approach to genome variant annotation, ClinVar may become a more robust and extensive knowledge base than any single locus-specific or laboratory-managed knowledge bases. Open access, locus-specific knowledge bases tend to be curated and

maintained on a volunteer basis, making the knowledge available limited. While laboratory-managed knowledge bases contain the best variant knowledge, they are also (1) limited by the number of unique variants observed by that laboratory and (2) may have tightly controlled access to the variant knowledge in order to maintain a competitive advantage over other testing laboratories [51]. Nevertheless, if ClinVar is embraced by the diagnostic laboratory community with the support of the ClinGen effort [52], the laboratory knowledge bases will likely serve as one of the most important sources of variant annotations.

Variant clinical interpretations categories

The clinical interpretation categories for sequence variations stored in the genome variant knowledge bases may follow recommendations set by the American College of Medical Genetics and Genomics (ACMG) and others [53,54]. The ACMG recommendations include classifications such as 'pathogenic,' 'likely pathogenic,' 'variant of unknown significance (VUS),' 'likely benign,' and 'benign' for diseases caused by genes. Pharmacogenomics (PGx) classification categories include such as 'ultrarapid metabolizer,' 'intermediate metabolizer,' and 'poor metabolizer' for genes impacting drug metabolism [55]. Some have also used allele classifications (e.g., *1/*2) to represent PGx variants for CDS, though this practice is becoming increasingly complicated as more variants are discovered [56]. Unfortunately, many of these classification categories are not used consistently and many labs create their own classification categories, which may result in interpretation discrepancies and confusion among clinicians. Therefore, in order to facilitate widespread adoption of CDS, future effort may be necessary to promote

standardized variant classification definitions [57].

Variant knowledge management

Our understanding of genomics in health is still relatively nascent. However, as research into the human genome grows, so too will understanding of health impacts of genome variants. This growth in understanding will likely lead to frequent and significant changes to variant classifications. To illustrate, over a seven year period, the Partners HealthCare Center for Personalized Genetic Medicine's Laboratory for Molecular Medicine genome variant knowledge base, managed by the GeneInsight Suite, reclassified nearly 15% of their original classifications, with almost one-third of those initially being VUS [30]. As such, genome variant knowledge bases will play an important role in independently managing the clinical interpretations of variants for the genomic CDS architecture. Not only can the most up-to-date variant classification be available during the annotation process, but if a clinical interpretation of a variant later changes, the variant classification for a particular patient's genome can be automatically updated. In such cases, changes in clinical interpretations will likely need to be versioned and tracked to account for potential liability concerns. Nevertheless, this separation of concerns through the SOA allows CDS to use the most up-to-date variant knowledge, while being free of dependencies that are timely and costly to update.

Genome databases

The storage of a patient's annotated genomic sequence is central to the proposed CDS architecture. With a patient's genome stored and accessible, a patient's genetic

information can be available for CDS when needed.

Genome data considerations

Although the size of a genome data set can be significantly reduced using variant file formats, much of the resulting data may still be unnecessary for most CDS use cases [9]. For example, these genome variant files contain a comprehensive set of all variants in the patient's genome, whether or not they are associated with a known gene or phenotype. As such, it may be unnecessary to make all variants available for CDS, particularly those which have no known association with genes and or phenotypic impact. Furthermore, while genome sequence metadata and annotations are important for quality assurance, variant classification, and versioning, some of this metadata may not be necessary for the purposes of CDS. Examples of this metadata include the reference sequence used, sequence coverage, population frequency, and reference copy number.

These examples are important to consider when trying to simplify CDS knowledge to the extent possible. To illustrate, in cases where there are hundreds of known pathogenic variants within a particular gene, it may not be efficient to write CDS knowledge for every known variant, particularly when the clinical phenotypes of different variants are identical and variant clinical interpretations can change. In certain use cases it may be sufficient to simply represent a variant by its clinical interpretations. For example, a simple CDS rule using this approach could be: "If [gene='MLH1'] has [variant classification='pathogenic'], then [recommendation='recommend colonoscopy to patient']." Nevertheless, for specific use cases where the variant location and effect (e.g., frameshift mutations) within a gene produces a unique phenotype, or when a particular allelic variant is important for

pharmacogenomic dosing, such information can still be made available to CDS when needed.

Database approach

As a result, for the purposes of CDS we advocate the use of a clinical genome database consisting of only a patient's clinically relevant variants and a full genome database consisting of all variants and genome metadata. The clinical genome database should consist primarily of variants in or near genome regions associated with phenotype (e.g., genes), with associated data elements required for CDS knowledge. Data elements and possible standards which could be used for CDS include (1) genome type (e.g., germline or somatic); (2) gene name in the standard Human Genome Organisation (HUGO) Gene Nomenclature Committee (HGNC) format [58]; (3) variant in a standard format, such as the Human Gene Variation Society (HGVS) format [59]; and (4) variant clinical classification, as provided by the genome variant knowledge base. Other potentially important elements that could be useful for CDS include genotype, haplotype, tissue type, and genome copy number. However, it is currently unknown exactly which genomic information will be necessary for CDS; future research will help determine which information is important.

Other reasons for creating a simplified clinical genome database are to improve performance and security. In a SOA architecture with multiple independent components, speed and efficiency are a top priority, particularly for CDS. Reducing the need for a database query to filter through unneeded data is likely to improve performance, particularly when such databases grow to include genomes of many patients. Furthermore,

limiting genetic information available to external queries promotes privacy and security, as clinically unnecessary genomic data could potentially be used to uniquely identify anonymous genomes [60].

Finally, while we describe the two databases as being separate, this can be a virtual separation or a physical separation. Nevertheless, there will need to remain a connection that will allow for changes in our understanding of the human genome. Indeed, data available in the full genome database will be available to the clinical genome database if and when they become clinically relevant and useful to CDS. Furthermore, just as a genome database is made available for clinical care, it should also be available for research, using many of the same service-based approaches [21].

The roles of the electronic health record

The EHR represents an important role in the proposed architecture, as it is responsible for collecting and storing the patient's clinical data required for CDS. Furthermore, it provides the mechanism by which CDS interacts with the end-user at the point and time of care. Indeed, to be effective, CDS for WGS should be integrated within the EHR clinical workflow, similar to how other nongenomic CDS is provided. It will likely not be sufficient or desirable to have a stand-alone CDS application for WGS information.

EHR as a repository of clinical data

EHRs serve as the primary source of collecting and storing clinical information that will be used to provide CDS. While EHRs have traditionally functioned as clinical data

repositories, most EHRs currently do not have an effective way of storing genetic information [61]. Furthermore, with competing higher-priority demands (e.g., Meaningful Use) among EHR vendors, this may not change in the near future. Therefore, our approach is to store genomic data separately from the clinical data in the EHR, and leverage service-based capabilities to obtain the clinical and genomic data required for CDS. This approach reduces the burden on EHR developers to build genome-specific capabilities, while allowing them to continue serving as the primary source of clinical data. Unfortunately, most EHRs use their own approaches for collecting and storing clinical data, which can be challenging for scalable CDS solutions [62]. However, this challenge can be overcome by mapping various data models to a common standardized data model, used specifically for CDS. A CDS data model being considered for EHR certification criteria related to Meaningful Use Stage 3 is the Health Level 7 Virtual Medical Record (vMR) standard [63].

CDS interface with end users

In addition to collecting and storing clinical data, EHRs are also responsible for the triggering of a CDS request and then presenting the CDS results in an effective way to end users. CDS can be triggered in a variety of situations, such as when (1) the patient's record is opened or a certain EHR view is selected, (2) a drug or procedure is ordered, (3) clinical documentation occurs within the EHR, or (4) at a routine time interval. Furthermore, the EHR can present CDS results within the clinical workflow of the clinician [14]. To this end, CDS results can be displayed as point of care alerts or reminders, relevant information displays, care recommendations, order facilitators, or workflow support [12,64]. In

principle, all the same CDS capabilities, which are currently available within EHRs, should also be used to trigger and present CDS for WGS information according to CDS best practices [13,14].

Leveraging available EHR capabilities

To provide CDS for WGS information within the EHR, the proposed architecture should primarily rely on EHR capabilities that are currently supported, or likely to be supported in the near future. To illustrate, the EHR market currently consists of hundreds of vendor solutions, each with their own development roadmaps and timelines [65]. Being reliant on custom EHR integration solutions may be an inadequate approach to attaining widespread and consistent use of CDS for WGS information [35]. Rather, aligning the proposed CDS architecture with current and potential future EHR capabilities, mandated by certification criteria, offers a pragmatic and effective solution. Of note, service-based CDS capabilities for EHRs are currently under consideration for Meaningful Use Stage 3 [25]. Moreover, some major EHR system vendors already support service-based CDS capabilities [66].

In summary, there are many advantages to leveraging EHR capabilities that are currently available and/or are aligned with relevant EHR certification criteria for WGS-driven CDS. As this approach is not dependent upon internal EHR development timelines and prioritization, it offers a greater chance of gaining widespread and consistent distribution across multiple EHR vendors and healthcare organizations. As such, this proposed architecture is designed to leverage existing EHR capabilities and align with ongoing developments in health IT.

CDS knowledge base

CDS entails providing person-specific care recommendations or knowledge, which can be used to enhance health and healthcare [67]. CDS knowledge bases contain representations of clinical knowledge (CDS knowledge) in the form of logic, decision rules, expressions, guidelines, and algorithms which support the provision of care, based upon a patient's clinical and genomic information. In a SOA CDS architecture, the CDS knowledge base is encapsulated as an independent unit by a service. This service receives patient-specific information provided by the CDS requester, processes this information, and returns a CDS result. As such, this approach reduces dependencies upon requesting EHR systems (requestor), if standardized data models and terminologies are used [68]. Furthermore, as the CDS knowledge is agnostic to how or where the data are originally stored, its primary concern is to process the standardized patient data according to the knowledge it contains.

Likewise, as CDS knowledge authoring and maintenance can be time consuming, keeping the maintenance of variant classifications separate from CDS knowledge will promote efficiency in CDS knowledge management. This approach allows variant classifications to freely change without needing to update CDS knowledge bases as well. Finally, as CDS knowledge could become complicated for genomic information, simplifying the knowledge to the extent possible is a desirable attribute. As described previously, this can be achieved by writing CDS knowledge using a gene and an associated clinical interpretation. Creating CDS knowledge for every possible variant within a particular gene, for which there are thousands of variants known and potentially many more unknown, will be inefficient [9].

CDS knowledge development and management

As a result of the SOA approach, CDS knowledge bases can be deployed and maintained by an independent entity specializing in the development and management of CDS knowledge. For example, an entity that specializes in developing and optimizing pharmacogenomic dosing regimens can deploy their knowledge as a service-based CDS knowledge base, allowing subscribing organizations to leverage the most up-to-date knowledge, provided by that entity [69]. Likewise, medical societies, which develop disease-specific care guidelines and recommendations, can deploy their work as a CDS knowledge base and allow member institutions to utilize the care guidelines and recommendations in the form of CDS [70]. Furthermore, the ability to leverage independently developed CDS knowledge could increase a health care organization's access to CDS capabilities, and promote competition among CDS knowledge authors. Similarly, this SOA approach also supports the ability to share the same CDS knowledge among many health care organizations. This is important because it is unlikely that a single healthcare organization will be able to maintain all its own CDS knowledge for WGS information, particularly for small and rural healthcare organizations with limited genomics expertise [9].

CDS controller

As previously described, genomic data required by the CDS knowledge base will not be stored with the clinical data from the EHR. Rather, a patient's genomic information will be stored and maintained in a separate genome database [21]. With the separation of clinical data from genomic data as proposed in this CDS architecture, a component which

links and coordinates the other components of the architecture together, will be required. Indeed, this is the primary role of the CDS controller, which is to combine clinical data from the EHR with genetic data from the clinical genome database into a complete data package for the CDS knowledge base. The CDS controller compares the received patient data to the CDS knowledge data requirements, which can include required data elements and desirable formats. The CDS controller can also facilitate workflow-appropriate triggering, perform terminology mappings, exclude unneeded clinical data, request additional data from other sources, and enable end-user interaction, as necessary [26]. The functions of a CDS controller, in our proposed CDS architecture, consists of the following sequential steps:

1. The CDS controller obtains clinical data from the EHR in a standardized format (e.g., vMR), as a result of a CDS trigger within the EHR.
2. The patient data are compared with the data requirements for the requested CDS knowledge module. In the case of our architecture, the CDS controller will identify that the patient's genomic information required by the CDS knowledge is missing, and will make a request to the genome database for that information.
3. The CDS controller obtains the patient's genomic information from the clinical genome database, as specified by the CDS knowledge data requirements.
4. The CDS controller then merges the patient's genome information with the clinical information into a single vMR file.
5. The complete data package is subsequently transmitted to the CDS

knowledge base for evaluation.

6. After CDS evaluation, the CDS controller receives the CDS response from the CDS knowledge base. At this point, the CDS controller can then process the CDS responses with additional workflow requirements (e.g., human review and approval of CDS recommendation), if necessary.
7. The CDS response is relayed to the EHR for end-user presentation.

While the CDS controller is described as being a separate component in this architecture, it is certainly feasible for the CDS controller to be an embedded function within an EHR. Indeed, such a scenario is described in another manuscript [28].

Genome interpreter

The genome interpreter, while not directly involved with CDS as described above, may be an important component to clinicians who desire to manually review variants in a patient's genome. As CDS may not be able to represent every possible clinical scenario, the capacity to manually review variants, clinical impact, and relevant metadata about a patient's genome will be important. Examples of genome interpreters include those provided by commercial genome annotation companies [47–49]. While these solutions are available as stand-alone applications, ideally they should be made available to clinicians within their EHR.

Results

Achieving the EHR and CDS WGS desiderata

An objective of the proposed CDS architecture for WGS information is to satisfy the requirements in the technical desiderata [19,20]. Table 4.1 represents a summary of barriers to using WGS information in the EHR and CDS, the desiderata requirements that are designed to address the barrier, and a description of how our proposed architecture attempts to satisfy each requirement in order to overcome the barrier.

Discussion

Comparison of proposed architecture to prior work on CDS for genomics

As described in the introduction, there is a growing research base on CDS interventions for genomics [17]. Indeed, the design and capabilities of many of these CDS examples provide the basis for the conceptual approaches described in our proposed architecture. As described earlier, of the organizations described in the Tarcy-Hornoch *et al.* review [17], all organizations developed and managed their own genome annotation process, each developed custom genome variant knowledge bases, and most had primitive CDS capabilities, primarily limited to PDF reports. Furthermore, report generation was dependent upon local experts unique at each institution, an approach that is unlikely to be scalable.

As a noteworthy example, the GeneInsight Suite is a stand-alone, Web-based interface designed to manage and communicate genome variants and clinical interpretations between clinicians and laboratories [30,34,77]. The GeneInsight Suite is an

example of an application that can support genome variant knowledge base managed by a laboratory, and maintain current clinical interpretations in a genome database. While this application focuses on managing variant knowledge and communicating updates to clinician end-users, the knowledge communicated is largely limited to the patient's genomic information, variant clinical interpretation, and a generic variant report. Furthermore, as the application currently exists separately from the EHR, its ability to leverage clinical data and provide patient-specific CDS based on clinical and genomic data within the EHR workflow, is limited [9]. Indeed, tighter integration with the EHR and CDS is an important future effort acknowledged by the developers of the GeneInsight application [30].

Furthermore, several groups have implemented preemptive pharmacogenomics (PGx) CDS within EHRs, namely the PREDICT (Pharmacogenomic Resource for Enhanced Decisions in Care & Treatment) project at Vanderbilt in Nashville, Tennessee [35,78]; a group at St. Jude Children's Research Hospital in Memphis, Tennessee [79,80]; and the CLIPMERGE (Clinical Implementation of Personalized Medicine through Electronic Health Records and Genomics) project at Mt. Sinai Hospital in New York City [36]. The PREDICT project provides an example of active CDS for genotype information that is integrated within the clinical workflow of the EHR. Indeed, this CDS capability for PGx is built into the order entry component of Vanderbilt's homegrown EHR system. Furthermore, all genotype results for a patient are stored in database repository, separate from the EHR, with actionable genotype results and their interpretations stored as a laboratory result within the EHR. As the CDS for this project was developed and built into the EHR by an internal panel of experts, its scalability is limited beyond their own

institution. Furthermore, the CDS rules do not incorporate clinically relevant non-genomic information into the decision process [35,78]. The St. Jude Children's Hospital also takes an approach of storing genetic test results directly in the EHR (Cerner). The EHR uses its native CDS capabilities to provide alerts and recommendations, which are developed and maintained by the institution. Again, this approach will likely be challenging to scale beyond their institution and beyond the PGx use case. Finally, for the CLIPMERGE project, actionable PGx genetic test results derived from institution's research biobank (BioMe) are combined with relevant clinical information extracted from the institution's EHR (Epic) in an external CLIPMERGE database. An external CDS rules engine also processes the patient data from the database and returns the results back to the EHR in real-time. The CLIPMERGE approach uses a separation of components and is likely the closest example to the approach described in the current manuscript. However, with clinical data extracted from the EHR and stored in a separate database along with the genetic information, it is also unclear if this approach could support WGS information and whether it can be easily scalable [57].

Originality and uniqueness of the proposed architecture

In summary, these examples represent important contributions to CDS approaches for genomics. While not all these solutions are designed for WGS information, and some of these approaches would struggle to support WGS information, they contain important design approaches that can be implemented in a scalable architecture, able to support WGS information. Indeed, many of the design principles in these examples were a source of inspiration and adopted for our proposed CDS architecture. Indeed, we believe it will

require the coordination of several of these proven components to build a CDS architecture capable of effectively leveraging WGS information. As a result, we have proposed an architecture, which uses many of these proven design approaches, that is able to provide CDS for WGS information on a widespread scale. We believe that our proposed architecture approach, described in this manuscript, will be important to achieving this goal.

Barriers still to overcome

While the proposed architecture aims to overcome many barriers related to genetic information, there are still many barriers to overcome before this architecture can be realized on a widespread scale. For instance, our understanding of the human genome and thus the annotation process is still in the relatively early stages. In fact, the reference genome used during the annotation process will likely change in the future. Likewise, many caveats such as race and family health history must be considered for an accurate clinical interpretation of a variant. Furthermore, as described earlier, many variant classification categories are used to describe the clinical impact of variants. As these classifications may be used when authoring CDS knowledge, it is important that a standard, well defined variant classification system is consistently used to describe a variant's clinical impact. While ClinVar has a set of classifications that are currently used [81], they are probably not sufficient to represent clinical impact with the specificity needed. Furthermore, with regards to ClinVar, there may be situations that arise involving differing expert interpretations for the same variant. In such cases, the various interpretations will have to be harmonized in some way by ClinVar or a related entity.

With regards to CDS infrastructure, service-based CDS capabilities are still in the early stages of industry adoption and thus still fairly limited with regards to technical capabilities, available standards, and available CDS knowledge. Indeed, using the SOA CDS approach described herein will require that significant gaps in standards and technology be addressed. Furthermore, even with the technical capabilities in place, there are still many nontechnical issues for service-based CDS that will need to be overcome, such as legal uncertainties regarding medical liability and questions regarding the financial sustainability of a services-based approach to CDS delivery. While such issues are important and must be addressed to enable services-based CDS for WGS information, these issues are also of interest to, and being addressed by, the larger CDS community. For instance, the consistent and widespread adoption of a service-based CDS architecture may be greatly enhanced by related EHR certification criteria that are under consideration for Meaningful Use Stage 3, due out in 2017 [82]. Indeed, efforts are currently underway in the Health eDecisions initiative to develop and pilot standards that are being considered for this purpose [83]. Of note, however, regulations and EHR certification criteria related to Meaningful Use Stage 3 are still under development and are subject to change. Nevertheless, some major EHRs have already implemented, or have plans to implement, service-based CDS capabilities in the near future, irrespective of Meaningful Use requirements.

Current efforts and future direction

While this manuscript is largely theoretical, current efforts by the authors are underway to build and test a functional prototype of this system, with greater technical

details regarding specific specifications. Indeed, this prototype currently follows the methods described in this paper in an attempt to meet the requirements in the technical desiderata. Once demonstrated with a prototype, it would be appropriate to build out a more robust infrastructure and implement the architecture on a small scale within a clinical setting. Such an implementation could begin with single gene test results, and then move to more complex gene panels and whole genome sequences. Additionally, research and experience from these implementations may determine that performance issues and security may be less of a concern than previous thought. Moreover, as the architecture capabilities become available to more health care providers, it will become appropriate to develop genome-specific CDS knowledge.

Also, as mentioned in the section on genome storage, future research will be needed to determine which genomic information will be essential for CDS knowledge. Indeed, a systematic review and analysis of potential CDS knowledge for genomic information could help determine the most important elements for genome-based CDS. Furthermore, the current architecture is primarily focused on (1) simple kinds of genomic variation (e.g., SNP variants within genes) and (2) variants with known clinical impact. However, current and future genomic discoveries may uncover complex interactions, which may require additional architectural considerations and modifications in order to support CDS. Likewise, by incorporating variant prediction algorithms such as VAAST, CDS could also become more involved with the interpretation of novel variants [45,46]. As a result, we do not presume the currently proposed architecture to be the final solution for WGS-based CDS. Rather, the current architecture provides a foundation for future development and modifications as our understanding of the genome and health grows.

Conclusion

The availability of a patient's whole genome sequence has the potential to facilitate the practice of personalized health care in the clinic. While research efforts are producing significant discoveries in support of personalized medicine, many barriers exists which limit the effective utilization of these discoveries in a clinical setting. Such barriers include the complexity of genomic information, the changing nature of understanding of the genome, current result reporting methodologies, and limited availability of clinical genomics experts [9]. However, effectively designed CDS provided within the clinical workflow, offers a potential solution to support the effective clinical utilization of WGS information. Indeed, a well-coordinated, service-based CDS architecture represents a practical solution to provide WGS-enabled CDS at the point of care.

References

1. Report of the President's Council of Advisors on Science and Technology. *Priorities for Personalized Medicine*.; 2008.

2. Abrahams E, Ginsburg GS, Silver M. The Personalized Medicine Coalition: Goals and strategies. *Am J Pharmacogenomics*. 2005;5(6):345–55. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16336000.

3. Bonetta L. Whole-genome sequencing breaks the cost barrier. *Cell*. 2010:917–919.

4. Ashley EA, Butte AJ, Wheeler MT, *et al.* Clinical assessment incorporating a personal genome. *Lancet*. 2010;375(9725):1525–35. doi:10.1016/S0140-6736(10)60452-7.

5. Lupski JR, Reid JG, Gonzaga-Jauregui C, *et al.* Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med*. 2010;362(13):1181–91. doi:10.1056/NEJMoa0908094.

6. Rope AF, Wang K, Evjenth R, *et al.* Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. *Am J Hum Genet*. 2011;89(1):28–43. doi:10.1016/j.ajhg.2011.05.017.

7. Talkowski ME, Ordulu Z, Pillalamarri V, *et al.* Clinical diagnosis by whole-genome sequencing of a prenatal sample. *N Engl J Med*. 2012;367(23):2226–32. doi:10.1056/NEJMoa1208594.

8. Wetterstrand K. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). 2013. Available at: http://www.genome.gov/sequencingcosts/. Accessed February 6, 2013.

9. Welch BM, Kawamoto K. The need for clinical decision support integrated with the Electronic Health Record for the clinical application of whole genome sequencing information. *J Pers Med*. 2013;3:306–325. doi:10.3390/jpm20x000x.

10. Downing GJ, Boyle SN, Brinner KM, Osheroff JA. Information management to enable personalized medicine: Stakeholder roles in building clinical decision support. *BMC Med Inform Decis Mak*. 2009;9:44. doi:10.1186/1472-6947-9-44.

11. Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. *J Am Med Informatics Assoc JAMIA*. 2007;14(2):141–145. doi:10.1197/jamia.M2334.Introduction.

12. Wright A, Sittig DF, Ash JS, *et al.* Development and evaluation of a comprehensive clinical decision support taxonomy: Comparison of front-end tools in commercial and internally developed Electronic Health Record systems. *J Am Med Inform Assoc*. 2011;18(3):232–42. doi:10.1136/amiajnl-2011-000113.

13. Bates D, Kuperman G. Ten commandments for effective clinical decision support: Making the practice of evidence-based medicine a reality. *J Am Med Informatics Assoc JAMIA*. 2003;10(6):523–530. doi:10.1197/jamia.M1370.Although.

14. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. *BMJ*. 2005;330(7494):765. doi:10.1136/bmj.38398.500764.8F.

15. Welch BM, Kawamoto K. Clinical decision support for genetically guided personalized medicine: A systematic review. *J Am Med Inform Assoc*. 2012;20(2):388–400. doi:10.1136/amiajnl-2012-000892.

16. Drohan B, Ozanne EMM, Hughes KSS. Electronic health records and the management of women at high risk of hereditary breast and ovarian cancer. *Breast J*. 2009;15:S46–S55. doi:10.1111/j.1524-4741.2009.00796.x.

17. Tarczy-Hornoch P, Amendola L, Aronson SJ, *et al.* A survey of informatics approaches to whole-exome and whole-genome clinical reporting in the electronic health record. *Genet Med*. 2013;15(10):824–32. doi:10.1038/gim.2013.120.

18. Overby CL, Kohane I, Kannry JL, *et al.* Opportunities for genomic clinical decision support interventions. *Genet Med*. 2013;15(10):817–23. doi:10.1038/gim.2013.128.

19. Masys DR, Jarvik GP, Abernethy NF, *et al.* Technical desiderata for the integration of genomic data into Electronic Health Records. *J Biomed Inform*. 2012;45(3):419–22. doi:10.1016/j.jbi.2011.12.005.

20. Welch BM, Eilbeck K, Del Fiol G, Meyer L, Kawamoto K. Technical desiderata for the integration of genomic data with clinical decision support. *Submitt Publ*. 2014.

21. Starren J, Williams MS, Bottinger EP. Crossing the omic chasm: A time for omic ancillary systems. *JAMA*. 2013;309(12):1–2. doi:10.1001/jama.2013.1579.

22. Erl T. *Service-Oriented Architecture (SOA): Concepts, Technology, and Design*. Prentice Hall; 2005:1–792. Available at: http://www.amazon.com/Service-Oriented-Architecture-SOA-Concepts-Technology/dp/0131858580. Accessed December 3, 2013.

23. He H. What Is Service-Oriented Architecture? *XML.com*. 2003. Available at: http://www.xml.com/pub/a/ws/2003/09/30/soa.html. Accessed December 1, 2013.

24. Juneja G, Dournaee B, Natoli J, Birkel S. Improving performance of healthcare systems with service oriented architecture. *InfoQ*. 2008:1–15. Available at: http://www.infoq.com/articles/soa-healthcare.

25. Kawamoto K, Lobach DF. Design, implementation, use, and preliminary evaluation of SEBASTIAN, a standards-based Web service for clinical decision support. *AMIA Annu Symp Proc*. 2005;2005380–4.

26. Kawamoto K, Jacobs J, Welch BM, *et al.* Clinical information system services and capabilities desired for scalable, standards-based, service-oriented decision support: Consensus assessment of the Health Level 7 clinical decision support work group. *AMIA Annu Symp Proc*. 2012;2012:446–55.

27. Kawamoto K, Del Fiol G, Orton C, Lobach DF. System-agnostic clinical decision support services: Benefits and challenges for scalable decision support. *Open Med Inform J*. 2010;4:245–54. doi:10.2174/1874431101004010245.

28. Kawamoto K, Lobach D. Proposal for fulfilling strategic objectives of the US roadmap for national action on decision support through a service-oriented architecture leveraging HL7 services. *J Am Med Informatics Assoc*. 2007:146–155. doi:10.1197/jamia.M2298.Introduction.

29. Standards & Interoperability (S&I) Framework - Health eDecisions Homepage.

Available at: http://wiki.siframework.org/Health+eDecisions+Homepage. Accessed November 8, 2013.

30. Aronson SJ, Clark EH, Varugheese M, Baxter S, Babb LJ, Rehm HL. Communicating new knowledge on previously reported genetic variants. *Genet Med*. 2012;14(8):713–719. doi:10.1038/gim.2012.19.

31. Kawamoto K, Lobach DF, Willard HF, Ginsburg GS. A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. *BMC Med Inform Decis Mak*. 2009;9:17. doi:10.1186/1472-6947-9-17.

32. Kawamoto K. Integration of knowledge resources into applications to enable clinical decision support: Architectural considerations. In: Robert A. Greenes, MD PhD, ed. *Clinical decision support: the road ahead*. 1st ed. Elsevier; 2007:503–538.

33. Hamilton AB, Oishi S, Yano EM, Gammage CE, Marshall NJ, Scheuner MT. Factors influencing organizational adoption and implementation of clinical genetic services. *Genet Med*. 2013. doi:10.1038/gim.2013.101.

34. Wilcox AR, Neri PM, Volk L a, *et al.* A novel clinician interface to improve clinician access to up-to-date genetic results. *J Am Med Inform Assoc*. 2013:1–5. doi:10.1136/amiajnl-2013-001965.

35. Pulley JM, Denny JC, Peterson JF, *et al.* Operational implementation of prospective genotyping for personalized medicine: The design of the vanderbilt PREDICT project. *Clin Pharmacol Ther*. 2012;92(1):87–95. doi:10.1038/clpt.2011.371.

36. Gottesman O, Scott S, Ellis S. The CLIPMERGE PGx Program: Clinical implementation of personalized medicine through Electronic Health Records and genomics–pharmacogenomics. *Clin Pharmacology*. 2013;94(2):214–217. doi:10.1038/clpt.2013.72.The.

37. Danecek P, Auton A, Abecasis G, *et al.* The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8. doi:10.1093/bioinformatics/btr330.

38. Reese MG, Moore B, Batchelor C, *et al.* A standard variation file format for human genome sequences. *Genome Biol*. 2010;11(8):R88. doi:10.1186/gb-2010-11-8-r88.

39. Stenson PD, Mort M, Ball E V, *et al.* The Human Gene Mutation Database: 2008 update. *Genome Med*. 2009;1(1):13. doi:10.1186/gm13.

40. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of

Genetic Medicine, Johns Hopkins University (Baltimore, MD), February 8, 2013. World Wide Web URL: http://omim.org/.

41. National Center for Biotechnology Information. ClinVar. *US Natl Libr Med*. 2012. Available at: http://www.ncbi.nlm.nih.gov/clinvar/. Accessed February 8, 2013.

42. Fokkema I, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. LOVD v.2.0: The next generation in gene variant databases. *Hum Mutat*. 2011;32(5):557–63. doi:10.1002/humu.21438.

43. Adzhubei IA, Schmidt S, Peshkin L, *et al.* A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248–9. doi:10.1038/nmeth0410-248.

44. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812–4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=168916&tool=pmcentrez&rendertype=abstract. Accessed November 15, 2013.

45. Hu H, Huff CD, Moore B, Flygare S, Reese MG, Yandell M. VAAST 2.0: Improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution matrix. *Genet Epidemiol*. 2013;37(6):622–34. doi:10.1002/gepi.21743.

46. Singleton M, Jorde L, Yandell M. PHEVOR: Integration of VAAST with phenomizer and the gene ontology for accurate disease-gene identification using only a single affected exome. In: *American Society of Human Genetics*.; 2013. Available at: http://www.ashg.org/2013meeting/abstracts/fulltext/f130122223.htm.

47. Omicia. Omicia. 2013. Available at: http://www.omicia.com/. Accessed February 6, 2013.

48. SV Bio. SV Bio. 2013. Available at: http://svbio.com/. Accessed February 6, 2013.

49. Cypher Genomics. Available at: http://www.cyphergenomics.com/. Accessed December 1, 2013.

50. Wade N. The quest for the $1,000 human genome: DNA sequencing in the doctor's office? At birth? It may be coming closer. *N Y Times Web*. 2006:F1, F3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16874933. Accessed December 3, 2013.

51. Cook-Deegan R, Conley JM, Evans JP, Vorhaus D. The next controversy in genetic testing: Clinical data as trade secrets? *Eur J Hum Genet*. 2013;21(6):585–8. doi:10.1038/ejhg.2012.217.

52. Riggs ER, Wain KE, Riethmaier D, *et al.* Towards a universal clinical genomics database: The 2012 International Standards for Cytogenomic Arrays Consortium Meeting. *Hum Mutat*. 2013;34(6):915–9. doi:10.1002/humu.22306.

53. Richards CS, Bale S, Bellissimo DB, *et al.* ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med*. 2008;10(4):294–300. doi:10.1097/GIM.0b013e31816b5cae.

54. Plon S, Eccles D, Easton D. Sequence variant classification and reporting: Recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat*. 2008;29(11):1282–1291. doi:10.1002/humu.20880.Sequence.

55. Crews KR, Gaedigk A, Dunnenberger HM, *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for codeine therapy in the context of cytochrome P450 2D6 (CYP2D6) genotype. *Clin Pharmacol Ther*. 2012;91(2):321–6. doi:10.1038/clpt.2011.287.

56. The Human Cytochrome P450 (CYP) Allele Nomenclature Database. Available at: http://www.cypalleles.ki.se/. Accessed December 1, 2013.

57. Kho AN, Rasmussen L V, Connolly JJ, *et al.* Practical challenges in integrating genomic data into the electronic health record. *Genet Med*. 2013;15(10):772–8. doi:10.1038/gim.2013.131.

58. Gray KA, Daugherty LC, Gordon SM, Seal RL, Wright MW, Bruford EA. Genenames.org: The HGNC resources in 2013. *Nucleic Acids Res*. 2013;41(Database issue):D545–52. doi:10.1093/nar/gks1066.

59. Horaitis O, Cotton RGH. The challenge of documenting mutation across the genome: The human genome variation society approach. *Hum Mutat*. 2004;23(5):447–52. doi:10.1002/humu.20038.

60. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. *Science*. 2013;339(6117):321–4. doi:10.1126/science.1229566.

61. Hoffman MA, Williams MS. Electronic medical records and personalized medicine. *Hum Genet*. 2011;130(1):33–9. doi:10.1007/s00439-011-0992-y.

62. Huff S. Ontologies, vocabularies, and data models. In: Robert A. Greenes, MD PhD, ed. *Clinical decision support: the road ahead*. 1st ed. Elsevier; 2007: 307-324.

63. Kawamoto K, Del Fiol G, Strasberg HR, *et al.* Multi-national, multi-institutional analysis of clinical decision support data needs to inform development of the HL7

virtual medical record standard. *AMIA Annu Symp Proc*. 2010;2010:377–81.

64. Horsky J, Schiff GD, Johnston D, Mercincavage L, Bell D, Middleton B. Interface design principles for usable decision support: A targeted review of best practices for clinical prescribing interventions. *J Biomed Inform*. 2012. doi:10.1016/j.jbi.2012.09.002.

65. Zhang NJ, Seblega B, Wan T, Unruh L, Agiro A, Miao L. Health information technology adoption in U.S. Acute care hospitals. *J Med Syst*. 2013;37(2):9907. doi:10.1007/s10916-012-9907-2.

66. Zhang M, Velasco F, Musser R, Kawamoto K. Enabling cross-platform clinical decision support through Web-based decision support in commercial Electronic Health Record systems: Proposal and evaluation of initial prototype implementations [In press]. *AMIA Annu Symp Proc*. 2013.

67. Teich JM, Osheroff JA, Pifer EA, Sittig DF, Jenders RA. Clinical decision support in electronic prescribing: recommendations and an action plan: Report of the joint clinical decision support workgroup. *J Am Med Inform Assoc*. 12(4):365–76. doi:10.1197/jamia.M1822.

68. Kawamoto K, Del Fiol G, Lobach DF, Jenders RA. Standards for scalable clinical decision support: Need, current and emerging standards, gaps, and proposal for progress. *Open Med Inform J*. 2010;4(919):235–44. doi:10.2174/1874431101004010235.

69. Gage B, Eby C, Johnson J. Use of pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. *Clin Pharmacol Ther*. 2008;84(10):326–331. doi:10.1038/clpt.2008.10.Use.

70. Recommendations from the EGAPP Working Group: Genetic testing strategies in newly diagnosed individuals with colorectal cancer aimed at reducing morbidity and mortality from Lynch syndrome in relatives. *Genet Med*. 2009;11(1):35–41. doi:10.1097/GIM.0b013e31818fa2ff.

71. Schrijver I, Aziz N, Farkas DH, *et al.* Opportunities and challenges associated with clinical diagnostic genome sequencing: A report of the association for molecular pathology. *J Mol Diagnostics*. 2012;14(6):525–540. doi:10.1016/j.jmoldx.2012.04.006.

72. Ast G. The alternative genome. *Sci Am*. 2005;292(4):40–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15915813. Accessed January 23, 2014.

73. Gottesman O, Kuivaniemi H, Tromp G, *et al.* The Electronic Medical Records and Genomics (eMERGE) Network: Past, present, and future. *Genet Med*. 2013;15(10):761–71. doi:10.1038/gim.2013.72.

74. Imamura M, Maeda S. Genetics of type 2 diabetes: The GWAS era and future perspectives. *Endocr J*. 2011;58(9):723–39.

75. Cystic Fibrosis Mutation Database: Statistics. Available at: http://www.genet.sickkids.on.ca/StatisticsPage.html. Accessed February 19, 2013.

76. US Department of Health and Human Services Office of the Secretary. *Standards for Privacy of Individually Identifiable Health Information; Final Rule*.; 2000:45 CFR 164.502(b), 164.514(d).

77. Aronson S, Clark E, Babb L. The GeneInsight Suite: A platform to support laboratory and provider use of DNA based genetic testing. *Hum Mutat*. 2011;32(5):532–536. doi:10.1002/humu.21470.The.

78. Peterson JF, Bowton E, Field JR, *et al.* Electronic health record design and implementation for pharmacogenomics: A local perspective. *Genet Med*. 2013;15(10):833–41. doi:10.1038/gim.2013.109.

79. Bell GC, Crews KR, Wilkinson MR, *et al.* Development and use of active clinical decision support for preemptive pharmacogenomics. *J Am Med Inform Assoc*. 2013. doi:10.1136/amiajnl-2013-001993.

80. Hicks JK, Crews KR, Hoffman JM, *et al.* A clinician-driven automated system for integration of pharmacogenetic interpretations into an electronic medical record. *Clin Pharmacology.* 2012;92(5):563–566. doi:10.1038/clpt.2012.140.A.

81. ClinVar standard terms for clinical significance. Available at: ftp://ftp.ncbi.nlm.nih.gov/pub/GTR/standard_terms/Clinical_significance.txt. Accessed December 1, 2013.

82. Meaningful use criteria and how to attain meaningful use of EHRs. *HealthIT.gov*. Available at: http://www.healthit.gov/providers-professionals/how-attain-meaningful-use. Accessed December 1, 2013.

1. Maintain separation of primary molecular observations from the clinical interpretations of those data
2. Support lossless data compression from primary molecular observations to clinically manageable subsets
3. Maintain linkage of molecular observations to the laboratory methods used to generate them
4. Support compact representation of clinically actionable subsets for optimal performance
5. Simultaneously support human-viewable formats and machine-readable formats in order to facilitate implementation of decision support rules
6. Anticipate fundamental changes in the understanding of human molecular variation
7. Support both individual clinical care and discovery science

Figure 4.1: Desiderata for the integration of genomic data into EHRs described by Masys *et al.*

8. CDS knowledge must have the potential to incorporate multiple genes and clinical information.
9. Keep CDS knowledge separate from variant classification.
10. CDS knowledge must have the capacity to support multiple EHR platforms with various data representations with minimal modification.
11. Support a large number of gene variants while simplifying the CDS knowledge to the extent possible.
12. Leverage current and developing CDS and genomics infrastructure and standards.
13. Support a CDS knowledge base deployed at and developed by multiple independent organizations.
14. Access and transmit only the genomic information necessary for CDS.

Figure 4.2: Additional desiderata for the technical integration of WGS with CDS

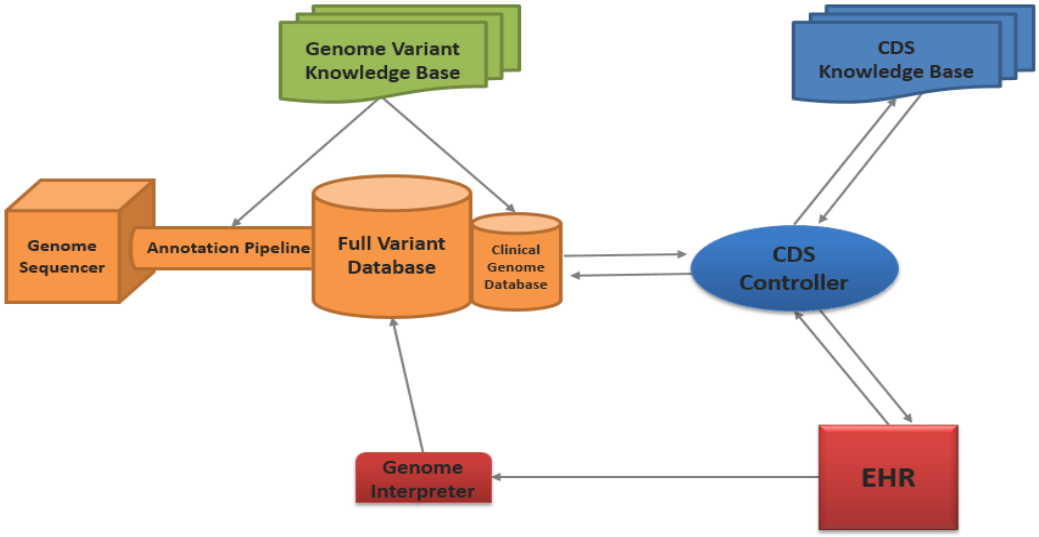


Figure 4.3: The proposed SOA architecture for WGS-enabled CDS

Table 4.1: A description of how the proposed architecture attempts to satisfy each desiderata requirement in order to overcome a WGS barrier to CDS

WGS barriers	Desiderata requirements	How the proposed architecture addresses requirements
Clinical interpretations of genomic information can be dynamic [30]	(Desiderata #1) Maintain separation of primary molecular observations from the clinical interpretations of those data	The *genome variant knowledge bases* exists separately and independently from the *genome databases*
WGS information contains a large amount of redundant and nonrelevant data [38]	(Desiderata #2) Support lossless data compression from primary molecular observations to clinically manageable subsets	Genome variant file formats are based on a reference sequence and a *clinical genome database* is used
Genomic results may be different based upon laboratory methods [71]	(Desiderata #3) Maintain linkage of molecular observations to the laboratory methods used to generate them	Laboratory methods are included with the variant file in the *full genome database*
A majority of a patient's 3,000,000+ genome variants will not have a clinical impact [4]	(Desiderata #4) Support compact representation of clinically actionable subsets for optimal performance	Compact representation of clinically actionable informatics are available in the *clinical genome database*
Computing on the genome will require data representations that are hard for humans to understand [61]	(Desiderata #5) Simultaneously support human-viewable formats and machine-readable formats in order to facilitate implementation of decision support rules	The machine-readable data format is used throughout the architecture, whereas a human viewable format is available through the *genome interpreter*
Our understanding of the human genome is nascent and in the future may change significantly [72]	(Desiderata #6) Anticipate fundamental changes in the understanding of human molecular variation	The proposed SOA architecture design allows for the flexibility of components to adapt to additional requirements as needed
Using available clinical and genomic information will be essential for research and discovery [73]	(Desiderata #7) Support both individual clinical care and discovery science	The same methods used to gather clinical and genomic data for CDS can be used for research as well
Relatively few diseases are caused by a single genetic variant alone [74]	(Desiderata #8) CDS knowledge must have the potential to incorporate multiple genes and clinical information	The *CDS controller* is able to collect all required clinical and genomic data required by the *CDS knowledge base*
CDS knowledge may evolve independently of variant classifications [30]	(Desiderata #9) Keep CDS knowledge separate from variant classification	The *CDS knowledge base* is a separate component from the *genome variant knowledge base*
Many organizations, with various EHR platforms, will likely not be able to develop their own CDS for WGS information [65]	(Desiderata #10) CDS knowledge must have the capacity to support multiple EHR platforms with various data representations with minimal modification	The architecture uses industry standards and approaches for scalable, interoperable CDS that are being considered for inclusion in EHR certification criteria related to Meaningful Use Stage 3
A single gene can have 100s-1000s of variants with various clinical impacts [75]	(Desiderata #11) Support a large number of gene variants while simplifying the CDS knowledge to the extent possible	The information in the *clinical genome database* and required for CDS can simply consist of the gene and its clinical interpretation
Re-inventing prior standards work on genomics and CDS just for this use case may prove to be futile [57]	(Desiderata #12) Leverage current and developing CDS and genomics infrastructure and standards	Health IT and genetics standards are used throughout the architecture where possible
No single entity will be able to develop and maintain all possible CDS knowledge for WGS [69]	(Desiderata #13) Support a *CDS knowledge base* deployed at and developed by multiple independent organizations	Service-based CDS supports CDS knowledge developed and maintained by multiple independent organizations
The file size and security concerns for WGS information are important [76]	(Desiderata #14) Access and transmit only the genomic information necessary for CDS	The *CDS controller* requests only the genome data needed for the CDS knowledge

CHAPTER 5

CLINICAL DECISION SUPPORT FOR WHOLE GENOME SEQUENCE INFORMATION LEVERAGING A SERVICE-ORIENTED ARCHITECTURE: A PROTOTYPE[3]

Introduction

Having a patient's whole genome sequence (WGS) information available to guide clinical decision-making has been an important goal of genome research, as WGS information could be used to improve clinical diagnosis, guide preventative efforts, and inform therapeutic decisions in the clinic [1]. Indeed, several clinical examples illustrate how WGS information has been used for the clinical diagnosis and treatment of patients with rare or previously undiagnosed diseases [2–4]. A patient's WGS information available at the point of care can increase a clinician's capacity to practice personalized medicine in a routine clinical care setting [5]. In fact, as a result of increasing availability of sequencing technology, as well as the exponentially declining costs of obtaining one's genomic information, the application of WGS information to routine clinical care is becoming increasingly possible [6].

[3] This chapter has been submitted as a conference paper for the 2014 American Medical Informatics Association Annual Symposium. Co-authors include Salvador Rodriguez-Loya, Dr. Karen Eilbeck, and Dr. Kensaku Kawamoto.

The need for clinical decision support for WGS

Although WGS information has the potential to be a valuable resource for clinical decision-making, its effective application to routine clinical care will likely be hindered by a number of challenges. Examples of such challenges include (1) static laboratory reports intended for human consumption, (2) the complexity of genetic analysis, (3) limited physician proficiency in genetics, and (4) the lack of genetics professionals in the clinical workforce [7]. Nevertheless, these challenges could be overcome by means of clinical decision support (CDS). CDS entails providing clinicians, patients, and other healthcare stakeholders with pertinent knowledge and/or person-specific information, intelligently filtered or presented at appropriate times, to enhance health and health care [8]. CDS provided within the clinical workflow, and at the point and time of decision-making within the electronic health record (EHR), has been shown to be the most effective way to deliver CDS to clinicians [9]. To realize the potential of genome-guided clinical care, CDS, leveraging a patient's WGS information, must be provided in real-time within the clinical workflow and the EHR [7].

Desiderata for integrating genomic information with the EHR and CDS

To address the complexity and challenges of integrating genomic information with EHRs, Masys *et al.* developed a set of desired technical requirements for EHRs to support the integration of genomic information [10]. See Figure 5.1 for the list of these requirements. Welch *et al.* developed an additional set of requirements, extending the Masys *et al.* desiderata, specifically addressing the technical requirements related to CDS

for WGS information [11]. See Figure 5.2 for a list of these requirements. Both of these efforts are aimed at providing guidance to system developers who are developing health IT capabilities for WGS information.

Proposed CDS architecture for WGS information

To satisfy the desiderata requirements described above, we previously proposed the use of a service-oriented architecture (SOA) to provide automatic CDS for WGS information at the point of care [12]. SOA is a software design approach which uses separate, independent software components known as services, which are self-contained components that have well-defined, understood capabilities [13]. SOA facilitates the reusability and standardization of processes, allowing for independent evolution and modifications to a particular service, reducing the burden of change on the overall system. Indeed, SOA offers many advantages to health information technology and CDS; as a result, its use is growing in health care [14]. Figure 5.4 shows a diagram of the SOA architecture proposed and described in further detail by Welch *et al.* [12]. A brief summary of each component is provided here:

- *Genome sequencing and annotation*- This component of the architecture is responsible for making the patient's genome information available and providing the genome with relevant annotations (e.g., location of variants relative to genes, impact of variants on genes, etc.). The annotation process leverages variant knowledge contained in the genome variant knowledge base.
- *Genome variant knowledge base*- This is a knowledge repository of known genome variants and assigned clinical interpretations. The primary

responsibility of these knowledge repositories is to maintain the most up-to-date and accurate variant interpretations, which is important as variant interpretations are known to change over time [15].

- *Genome database*- In the proposed architecture, a patient's genome is stored in a genome database, separate from clinical information (e.g., EHR). This database securely stores patients' genomes along with the most up-to-date variant interpretations from the genome variant knowledge base. This database also provides a standardized interface for access to a patient's genomic information.
- *CDS knowledge base*- The role of the CDS knowledge base is to process the patient-specific clinical and genomic information provided to it. It subsequently returns patient-specific, knowledge-based recommendations and/or information to support clinical care.
- *Electronic health record*- The EHR provides the patient's clinical information used by the CDS knowledge base. The EHR is also responsible for actions which trigger CDS requests and the presentation of CDS results within the clinical workflow.
- *CDS controller*- The CDS controller is responsible for processing a CDS request and assuring that all required data for CDS knowledge are available. If additional patient data are needed for CDS, the CDS controller makes a request to other data repositories (e.g., genome database) for the required information.

This manuscript describes our effort to build a functional prototype of the previously proposed CDS architecture for WGS information [12]. In this paper, we

describe the open-source components and health IT standards used to develop this prototype. Furthermore, we describe the methods used to evaluate the prototype using a clinical use case involving a patient at increased risk for Lynch syndrome. Finally, we have identified areas that will require additional research and development in the future. While others have built CDS capabilities for genomics [16–18], to our knowledge this work is the first to describe a system that adheres to the technical desiderata [10,11] and uses a SOA approach to provide WGS-guided CDS within the clinical workflow within the EHR.

Materials and Methods

Components used and configuration

Genome data acquisition and database storage

We used the 10Gen data set, which represents the first ten publicly available human genomes in a standardized genome variation format (GVF), for the genomic information used in this study [19]. These genomes represent three different ethnicities (African, Asian, and Caucasian) and several sequencing platforms including SOLiD, Illumina, Sanger, Roche 454, CGenomics, and Helicos [20]. The genomes were annotated using the Web-based Omicia Opal genome annotation platform.

Genome database

To store a patient's genomic information and make it available for CDS data requests, we implemented a relational database using MySQL Community Server (version 5.6.15) and the HeidiSQL open-source database manager [21]. In this database, we created a table named 'patient_genome' consisting of seven columns (with column names in

quotes): (1) ‘MRN’ which is the patient’s unique medical record number (MRN) that matches the patient’s MRN in the Tolven open-source EHR; (2) ‘gene’ is the gene where the variant resides, represented using HUGO (Human Genome Organization) Gene Nomenclature Committee (HGNC) standardized nomenclature; (3) ‘refSNP’ is the reference SNP ID number; (4)‘nuc_var’ is the nucleotide variant represented in Human Genome Variation Society (HGVS) nomenclature; (5) ‘pro_var’ is the protein variant also represented in HGVS nomenclature; (6) ‘interpretation’ is the clinical impact of the variant provided by the ClinVar genome variant knowledge base; and (7) ‘id’ is an auto-incremented value for the primary key of the table. The 10Gen genomes were exported from Omicia as CSV files and then imported into the database through a database import process available through HeidiSQL. Code to regenerate the tables and import the genomes is available in Appendix D.

To support external access to a patient’s genomic data stored in the database, we developed a Web service interface deployed as a Java application within the JBoss Enterprise Application Platform version 6.1. This Web service interface provides access to the database content using the HL7 Retrieve, Locate, and Update Service (RLUS) standard [22]. The RLUS specification can be adapted to different semantic content formats. For the present prototype, we used the HL7 Virtual Medical Record (vMR) as the semantic content format within the RLUS-based Web service. See http://tinyurl.com/n3zksjd for the location of the RLUS specification used in this prototype. The location of the code used for the Web service interface can be found at https://bitbucket.org/Salvador_Rdz/genome-data-app-hl7-omg-rlus-interface.

ClinVar genome variant knowledge base

We used ClinVar as the genome variant knowledge base in this prototype. ClinVar is a publicly available repository of human genome sequence variations and associated phenotypes supported by the National Center for Biotechnology Information of the U.S. National Library of Medicine [23]. ClinVar currently does not support service calls to its knowledge base, so we replicated this functionality by developing a second table in the genome database called 'genomekb' and imported a subset of ClinVar data into this table. A ClinVar full data release was downloaded directly from the ClinVar FTP site to a Linux server and transformed into a format suitable for database import using XSLT [24]. See Appendix D for the code used to download and transform the ClinVar file. We created a SQL statement to update the interpretation field in the 'patient_genome' table from the interpretation field in the 'genomekb' table, based upon matches in the gene and variant between the two tables. This SQL code is also available in Appendix D.

Tolven electronic health record

To store the patient's clinical information and provide a clinical workflow interface for CDS, we used the open-source Tolven electronic Clinician Health Record (eCHR™), which is provided as part of the Tolven platform [25]. Tolven eCHR™ is an Office of the National Coordinator (ONC) certified clinical information system which supports basic clinical processes and information exchange. The Tolven platform supports several additional components including ePrescribing, scheduling, and analytics. Tolven was selected because of (1) its open-source code; (2) its ability to be configured and customized; (3) its data model, which is based on the HL7 Reference Information Model;

and (4) its use of several terminology standards such as Current Procedural Terminology (CPT), Logical Observation Identifiers Names and Codes (LOINC), and RxNorm. Furthermore, third-party plugins can be developed to extend the functionality of the Tolven platform. We developed a plugin for Tolven using Java Enterprise Edition (version 1.6). The plugin is designed to serve three important purposes for the prototype: (1) to access the patient's clinical data stored in Tolven; (2) to create a vMR document containing the patient's data; and (3) to communicate with an external CDS Web service using the HL7 (Decision Support Service) DSS standard.

SwitchYard CDS controller

The CDS controller plays a central coordinating role in this architecture. It first processes a CDS request from Tolven and checks that all required data for CDS inferencing are available. It identifies that the genomic data are not available so it makes a data request to the genome database for additional information. We used the open source solution SwitchYard (JBoss) to provide these functions. SwitchYard is a component-based SOA development framework which combines several useful components and functions for SOA into one application, including Apache Camel, Java Enterprise Edition, business process management (BPM), orchestration, routing, validation, and transformation. Within SwitchYard, we developed a composite service that includes five components: CamelServiceRoute, ProcessComponent, VerifyGenomeData, RequestGenomeData and IntegrateGenomeData. Figure 5.5 represents the workflow of the VerifyGenomeData component which checks for genome data in the vMR and, if it is not available, makes a request to the genome database for the requested information. We also configured two Web

service adaptors for SwitchYard to request and receive data from other components of the CDS architecture. These include (1) an adaptor for the HL7 RLUS standard to interface with the genome database Web service and (2) an adaptor for the HL7 DSS standard to interface with the Tolven EHR and OpenCDS.

OpenCDS

For the final component of the architecture, we used OpenCDS (http://www.OpenCDS.org) to serve as the CDS knowledge base. OpenCDS is an open-source, Java-based CDS framework designed to enable the delivery of CDS as a Web service compliant with the HL7 DSS standard. OpenCDS provides a knowledge authoring environment and CDS rules engine. Internally, OpenCDS uses the vMR as its data model and JBoss Drools as its inferencing engine. OpenCDS also manages and uses its own terminology within the Apelon Distributed Terminology System (DTS), with mappings created to standard terminologies such as the Systematized Nomenclature of Medicine — Clinical Terms (SNOMED CT) and LOINC. For this prototype, we deployed a new instance of OpenCDS and created the CDS rule logic in the Web-based knowledge authoring environment for Drools known as Guvnor. Figure 5.6 shows a CDS business rule authored in Guvnor and used in this prototype. After the CDS rule was created and tested in Guvnor, it was deployed within the OpenCDS run-time environment. Several vocabulary terms required by the rules were added to the OpenCDS terminology in the Apelon DTS instance used by OpenCDS. The vocabulary terms and location of the knowledge developed for this prototype can be found in Appendix D.

Standards used

HL7 Decision Support Service standard

The DSS standard specifies a standard interface for providing CDS as a service [26] and is adopted by both HL7 and the Object Management Group (OMG) standards development organizations. The DSS standard includes three major interfaces including (1) evaluation, used to evaluate patient data and generate patient-specific results; (2) metadata discovery, to identify metadata and knowledge modules of a service; and (3) query, which is used to query for knowledge modules of interest. For this prototype, we used the evaluateAtSpecifiedTime request interface, which includes a payload section in which a base64 encoded version of the vMR is placed.

HL7 Virtual Medical Record Standard

The Virtual Medical Record (vMR) is a standard HL7 clinical data model designed for CDS [27]. The vMR data model can represent patient-specific classes such as demographics, encounters, procedures, problems, medications, laboratory results, and observations. A patient's clinical data are modeled using these various vMR classes and elements. To represent the patient's genome information, we used the ObservationResult class. Specifically, observationFocus was used to represent the gene name in HGNC format, observationValue was used to represent the nucleotide variant in HGVS format, and interpretation was used to represent the variant clinical interpretation in LOINC. See Figure 5.7 for an example genome variant result in vMR format. The most recent HL7-approved version of the vMR is Release 2. However, for this prototype we used Release 1 due to the availability of support for this release in OpenCDS.

HL7 Retrieve, Locate, and Update Service

The HL7 Retrieve, Locate, and Update Service (RLUS) defines the service interface to locate, retrieve, and update resources among and within healthcare organizations [22]. This specification was designed to support SOA in healthcare. The HL7 RLUS specification defines several methods including *describe*, *discard*, *get*, *initialize*, *list*, *locate*, and *put*. In our prototype, we used RLUS to request a patient's genomic information from the genome database using the *describe* and *get* methods. *Describe* returns a detailed schema definition and *get* retrieves patient data based on parameters supplied in the RLUS retrieval request.

Gene and variant nomenclature

Several genome standards were used to represent genomic information in this prototype. To represent genes, we used names approved by HGNC [28]. We used three types of standardized genome variant representations in the prototype: (1) the refSNP number assigned to a sequence variant by dbSNP [29]; (2) the nucleotide variant in HGVS nomenclature format for coding sequence [30]; and (3) the protein variant in HGVS nomenclature for protein variation. Finally, we used LOINC to represent the possible interpretations of gene variants, which include pathogenic, presumed pathogenic, unknown significance, benign, and presumed benign [31].

The CDS process

The previous sections describe the components used in the prototype and their configurations. This section describes the overall process for providing patient-specific

CDS using WGS information within the EHR. With all components of the prototype in place and functioning properly, the sequential steps of the application are as follows:

1. A patient's clinical data are recorded in Tolven by the clinician. When the patient's chart is modified and saved, a CDS request is triggered by the OpenCDS plugin in Tolven. The plugin proceeds to gather the patient's clinical data from Tolven and transform the data into the vMR format. When the patient's clinical data are all in vMR format, they are base64-encoded, placed into the DSS payload, and sent to the SwitchYard CDS controller.
2. When SwitchYard receives the DSS request with clinical data in vMR format, it validates the data against a vMR schema template defining the required data for the requested CDS knowledge module. In this case, SwitchYard identifies that the genomic data required by the CDS module are missing. It then creates a RLUS request to obtain that required information from the genome database.
3. The genome database Web service interface receives the RLUS request from SwitchYard. This interface retrieves the requested information from the genome database, which includes the gene and clinical interpretation (the return of specific variants are also available upon request). The interface creates a response RLUS message and inserts a vMR payload of the genomic information.
4. SwitchYard receives the RLUS response from the genome Web service interface and merges the genome vMR with the clinical data vMR from Tolven. If all the required data are present, a DSS request with the newly

merged vMR is sent to OpenCDS.

5. OpenCDS receives the DSS request and processes the data contained in the vMR against rules in the requested knowledge module (*MLH1*). A CDS result is produced and sent back to SwitchYard in a DSS response.
6. SwitchYard receives the response from OpenCDS and forwards the message to Tolven.
7. When the Tolven plugin receives the response, it renders the response as an alert within the Tolven user interface such that the CDS is provided within the clinical workflow and at the time of decision-making,

Clinical use case and evaluation methods

To assess whether the prototype is functional we implemented the following clinical scenario and evaluated the performance of the prototype.

Clinical use case: Lynch syndrome risk assessment

Lynch syndrome (or hereditary nonpolyposis colorectal cancer) is an autosomal dominant genetic condition caused by pathogenic mutations in mismatch repair genes, such as *MSH2*, *MLH1*, *MSH6*, and *PMS2* [32]. These mutations impact the ability of a cell to repair DNA replication errors that occur during cell division. As the cell continues to divide, these mutations continue to accumulate throughout the genome. These mutations may ultimately lead to uncontrolled cell proliferation and thus cancer. Patients who have a pathogenic variant in any of these genes possess the greatest risk for colorectal cancer, and they are also at increased risk for cancer in the stomach, intestines, liver, brain, skin and

other body sites. Every year in the U.S., approximately 4,000-7,000 new cases of colorectal cancer are caused by Lynch syndrome [32]. It is recommended that patients at increased risk for Lynch syndrome should receive a colonoscopy every one to two years starting around the age of 20 years [33]. While family history is a strong indicator for risk, genetic testing is the definitive test for a patient's Lynch syndrome risk. Genetic screening for the disease, although currently expensive, will likely become easier when WGS information is clinically available for routine care [34]. Even then, some clinicians may not be familiar with Lynch syndrome nor the best practices for risk mitigation and management. CDS can play an important role in the awareness and management of Lynch syndrome.

Implementation of the clinical use case

To assess the ability of this prototype architecture to provide point-of-care CDS for Lynch syndrome, we created and implemented the example clinical scenario found in Figure 5.3. To implement this clinical scenario in the prototype, we created a new 33 year old patient in Tolven. For simplicity and demonstration purposes, we tested the architecture using a single gene (*MLH1*) with a pathogenic variant. We assigned the publicly available Venter genome to the test patient, which did not have a pathogenic mutation in the *MLH1* gene [35]. Therefore, we created a known pathogenic *MLH1* mutation using a SQL script described in Appendix D. This genome change was created previous to a genome knowledge base update of the 'interpretation' database field in the patient_genome table. A CDS business rule representing the *MLH1* recommendation for colonoscopy was created in Guvnor and deployed on the OpenCDS run time environment (see Figure 5.6). The CDS query is triggered automatically each time the patient's record is modified.

Performance evaluation

To test the performance of the architecture we used the free, open-source load testing software LoadUI (www.loadui.org) running on a Windows 7 (64-bit) machine with two processors and four gigabytes of RAM. In order to identify sources of latency, we tested each service component of the architecture separately, which included the genome application service running on a Linux Ubuntu 12.04 (64-bit) server with eight processors and 16 gigabytes of RAM, as well as the OpenCDS service running on a Linux CentOS v5.8 (64-bit) with four processors and four gigabytes of RAM. We also tested the performance of the SwitchYard CDS Controller (also running on the Linux Ubuntu server previously described) which included service calls to both the genome application and OpenCDS within its evaluation. As such, the evaluation of the CDS controller most closely represents the overall performance of the architecture. Each component was tested with a random load balance of 100 simultaneous users every 10 seconds using the same data requirements described in the clinical scenario. The overall performance evaluation was limited to 20 simultaneous users as a result of performance issues related to running Switchyard on a machine with limited processor capabilities.

Results

The objective of this study was to develop a functional prototype of our proposed architecture [12]. To that end, we implemented a prototype and verified that a patient with a clinical scenario described in Figure 5.3 was evaluated properly by the prototype. Figure 5.8 shows a screenshot of the CDS recommendation successfully generated through the use of this prototype architecture within the Tolven EHR.

Architecture performance

The genome application (which includes the service interface and database request) handled 3,109 requests over a five minute period. This genome application's fastest request took 25 milliseconds (ms), and the slowest took 697 ms, with an average of 40 ms (SD 47.44 ms). OpenCDS handled 3,015 requests over a five minute period. The fastest request took 7 ms, and the slowest took 914 ms, with an average of 12 ms (SD 17.04 ms). We also evaluated the performance of the CDS controller, which includes service calls to the genome application and OpenCDS and thus closely represents the overall performance of the architecture. Due to hardware limitations of our machine, we limited the testing to 20 simultaneous users. The CDS controller handled 650 requests over a five minute period. The fastest request took 356 ms, the slowest took 4,243 ms, with an average of 944 ms (SD 621.04). It is important to note that the average response time was under one second. See Table 5.1 for a performance summary.

Discussion

As demonstrated by this prototype and demonstration, a service-oriented CDS architecture is able to support the provision of genome-guided CDS at the point of care within the EHR. Indeed, this prototype leverages many standards and open-source solutions, and it is capable of integrating with current health IT architectures and workflows.

Issues identified

Throughout the process of developing the prototype of this architecture, several issues were identified. We used ClinVar as the genome variant knowledge base for this prototype because it was easily accessible and publicly available. Unfortunately, ClinVar is still in a very early stage with regards to its variant knowledge base, limiting its ability to provide clinical interpretations for relevant variants. Specifically, the available variant knowledge in ClinVar is a small, but growing, subset of all available variant knowledge [36,37]. Nevertheless, it is expected that ClinVar will improve over time as more laboratories begin contributing their variant knowledge. Also, variants reported to ClinVar can range in quality on a scale from one to five stars, ranging from a single submitter's interpretation to variants reviewed and submitted by expert panels. This variable quality of variant interpretation or potential disagreement between submitters can be challenging for CDS to manage. Ideally, variant interpretations used in CDS should meet a minimum threshold of quality and confidence. A recent NIH-funded initiative called ClinGen is expected to improve the data available in ClinVar [38]. Although other genome variant knowledge bases are also available (such as the Human Genome Mutation Database), we did not integrate them with this prototype due to the time and cost associated with each integration. Integrating such variant knowledge bases into the architecture is an anticipated future effort. Nevertheless, once ClinVar becomes more developed, we believe it represents an ideal resource for genome variant knowledge management because of its public financing and its potential to become the largest single repository of genome variant knowledge [12]. Finally, many of the variant locations were based upon differing reference genomes, making it challenging to match variants in ClinVar to variants in the patient's

genome. Ideally, all variants used architecture should be updated to a single reference genome [39].

Another important issue we identified in this work is the speed and response time of the CDS architecture. The first of Bates *et al.* 10 Commandments for effective CDS is that 'Speed is everything,' in that CDS response time should be less than a second [40]. Unfortunately with SOA architecture, because several disparate components are involved to provide CDS, speed and performance can be a major challenge. This is indeed the case with our prototype. With heavy loads (>30 simultaneous users), we observed response times to increase dramatically to be several seconds. Moreover, as CDS knowledge and data requirements become more complex, this response time will likely increase. Although our results show that around 20-30 simultaneous requests is the most the prototype could process before it became unreasonable for CDS, many computer science techniques such as increasing processing power, parallel processing, and load balancing can be used to improve performance of the overall CDS processes. Nevertheless, the goal of this study was to show that the architectural approach could work. We believe this work has accomplished this goal, however, we still believe there is much testing and optimization that can be done to improve the performance and response time of the architecture.

Strengths and limitations

An important strength of this work is the use of freely available, open-source components. Indeed, anyone with sufficient training in the technologies used could rebuild this architecture without needing to purchase and use proprietary software or components. Another strength of the architecture is the use of available health IT standards wherever

possible. Such standards and terminologies include the HL7 DSS standard, HL7 vMR standard, LOINC, HGNC, and HGVS. As a result, we were able to demonstrate that standards-based approaches are feasible for increasing the likelihood of interoperability with other health IT systems in the future.

A limitation of the study is that we only demonstrated this architecture working using a simple scenario in one clinical use case (Lynch syndrome risk assessment). While, there are many other clinical and genomic scenarios (such as pharmacogenomics) that could be tested and demonstrated with this architecture, the scope of this effort was limited to assessing whether the proposed prototype could deliver standards-based CDS within the clinical workflow and within the EHR. Nevertheless, we are planning to demonstrate the extensibility of the architecture with various clinical genomic use cases in future work. Another limitation of the study is that we have only integrated the CDS architecture with a single EHR. While the primary aim of the current effort was to demonstrate minimum feasibility, future efforts will need to focus on integrating the architecture with other EHRs. Of note, the integration of SOA CDS capabilities into various EHRs is part of larger OpenCDS initiatives, outside the scope of the current research on CDS for WGS information.

Future direction

While we were able to demonstrate the feasibility of this architecture using open-source solutions and standards, it took a significant amount of configuration of these components to make this possible. As a result, a future direction of this research is to develop this CDS architecture into an easily deployable format for healthcare organizations

to set up and run with minimal modification. Such a solution could potentially consist of a virtual machine image which includes a preconfigured genome database already linked to several genome variant knowledge bases, OpenCDS configured for genome rule authoring, and SwitchYard set up to support WGS-enabled CDS. Furthermore, it will be important to develop and include integration plugins for commercial EHR systems. Indeed, EHR systems are starting to support service capabilities and future Meaningful Use guidelines may require all EHRs to support them, meaning this architecture has the potential to be used on a widespread basis for WGS CDS in the near future [41]. Likewise, as the standards-based CDS integration with EHRs become more widespread, it will be important to test the functionality with end-users and conduct clinical trials evaluating the clinical impact of genome-guided CDS using this approach.

Another important future effort is to build the genomic CDS knowledge base and expand the current genome database. As this architecture supports a CDS knowledge base that can serve multiple health care organizations, it may be possible to develop a genomic CDS knowledge base which could be shared among these organizations. Such an effort could involve the collaboration of informaticists, geneticists, and clinical experts to implement published clinical genomics guidelines into a computable CDS knowledge representation. Furthermore, such experts could also review the literature and current clinical practices to develop new genomic CDS knowledge for clinical care. During this processes of developing a genomic CDS knowledge base, additional data requirements for inclusion in the genome database will likely be identified. For example, in the current version of the genome database, only genes and variants are included. While these data may be sufficient for Lynch syndrome and other autosomal dominant use cases, these data

would not be sufficient for autosomal recessive genetic conditions and other types of genomic use cases (e.g., SNP-based testing). As a result, it will be necessary to add additional genomic information for CDS, such as chromosome number, zygosity, tandem repeat number, or sequencing quality scores. These additional genomic data requirements could potentially be identified though a systematic review of the literature and the involvement of various domain experts.

Conclusion

Using WGS information in the clinic for routine clinical care may be challenging for clinicians to manage without assistance. CDS provided within the clinical workflow, at the time of decision-making, provides a feasible solution to enable genetically-guided personalized medicine. To evaluate this potential solution, we developed and tested a functional CDS architecture for WGS information using SOA design principles. Through this effort, we were able to demonstrate that a functional prototype of this approach is capable of providing genome-guided CDS within the clinical workflow, and within the EHR. While future research and development is necessary before such an approach can be used in a clinical setting, this study demonstrates that the approach is feasible and valid. We therefore speculate that this work will help guide future research and development on the use of WGS-based CDS to support personalized healthcare.

References

1. Collins FS, Green ED, Guttmacher AE, *et al.* A vision for the future of genomics research. *Nature*. 2003;422:835–47. doi:10.1038/nature01626

2. Ashley EA, Butte AJ, Wheeler MT, *et al.* Clinical assessment incorporating a

personal genome. *Lancet.* 2010;375:1525–35. doi:10.1016/S0140-6736(10)60452-7

3. Talkowski ME, Ordulu Z, Pillalamarri V, *et al.* Clinical diagnosis by whole-genome sequencing of a prenatal sample. *N Engl J Med.* 2012;367:2226–32. doi:10.1056/NEJMoa1208594

4. Lupski JR, Reid JG, Gonzaga-Jauregui C, *et al.* Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med.* 2010;362:1181–91. doi:10.1056/NEJMoa0908094

5. Report of the President's Council of Advisors on Science and Technology. *Priorities for Personalized Medicine.* 2008.

6. Wetterstrand K. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). 2013. http://www.genome.gov/sequencingcosts/. Accessed 6 Feb 2013.

7. Welch BM, Kawamoto K. The need for clinical decision support integrated with the Electronic Health Record for the clinical application of whole genome sequencing information. *J Pers Med.* 2013;3:306–25. doi:10.3390/jpm20x000x

8. Osheroff JA, Teich JM, Middleton B, *et al.* A roadmap for national action on clinical decision support. *J Am Med Informatics Assoc.* 2007;14:141–5.

9. Kawamoto K, Houlihan CA, Balas EA, *et al.* Improving clinical practice using clinical decision support systems: A systematic review of trials to identify features critical to success. *BMJ.* 2005;330:765. doi:10.1136/bmj.38398.500764.8F

10. Masys DR, Jarvik GP, Abernethy NF, *et al.* Technical desiderata for the integration of genomic data into Electronic Health Records. *J Biomed Inform.* 2012;45:419–22. doi:10.1016/j.jbi.2011.12.005

11. Welch BM, Eilbeck K, Del Fiol G, *et al.* Technical desiderata for the integration of genomic data with clinical decision support [Under review]. 2014.

12. Welch BM, Loya SR, Eilbeck K, Kawamoto K. A proposed clinical decision support architecture capable of supporting whole genome sequence information. *Journal of Personalized Medicine*. 2014; 4(2):176-199.

13. Erl T. Service-Oriented Architecture (SOA): Concepts, Technology, and Design. Prentice Hall 2005.

14. Kawamoto K, Lobach DF. Design, implementation, use, and preliminary evaluation of SEBASTIAN, a standards-based Web service for clinical decision support. *AMIA Annu Symp Proc.* 2005;380–4.

15. Aronson SJ, Clark EH, Varugheese M, *et al.* Communicating new knowledge on previously reported genetic variants. *Genet Med*. 2012;14:713–9. doi:10.1038/gim.2012.19

16. Welch BM, Kawamoto K. Clinical decision support for genetically guided personalized medicine: A systematic review. *J Am Med Inform Assoc*. 2012;20:388–400. doi:10.1136/amiajnl-2012-000892

17. Tarczy-Hornoch P, Amendola L, Aronson SJ, *et al.* A survey of informatics approaches to whole-exome and whole-genome clinical reporting in the electronic health record. *Genet Med*. 2013;15:824–32. doi:10.1038/gim.2013.120

18. Overby CL, Kohane I, Kannry JL, *et al.* Opportunities for genomic clinical decision support interventions. *Genet Med*. 2013;15:817–23. doi:10.1038/gim.2013.128

19. Reese MG, Moore B, Batchelor C, *et al.* A standard variation file format for human genome sequences. *Genome Biol*. 2010;11:R88. doi:10.1186/gb-2010-11-8-r88

20. The Sequence Ontology - Resources - 10Gen Data Set. http://www.sequenceontology.org/resources/10Gen.html. Accessed 23 Jan 2014.

21. HeidiSQL - MySQL and MSSQL made easy. http://www.heidisql.com/. Accessed 23 Jan 2014.

22. HL7 Standards Product Brief - HL7 Version 3 Standard: Retrieve, Locate, and Update Service (RLUS) Release 1. http://www.hl7.org/implement/standards/product_brief.cfm?product_id=89. Accessed 3 Feb 2014.

23. National Center for Biotechnology Information. ClinVar. U.S. Natl. Libr. Med. 2012.http://www.ncbi.nlm.nih.gov/clinvar/. Accessed 8 Feb 2013.

24. ClinVar. ClinVarFullRelease_2014-01.xml.gz. 2014.ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/xml/. Accessed 1 Jul 2014.

25. Tolven Platform. http://home.tolven.org/. Accessed 26 Feb 2014.

26. HL7 Standards Product Brief - HL7 Decision Support Service (DSS). https://www.hl7.org/implement/standards/product_brief.cfm?product_id=12. Accessed 3 Feb 2014.

27. Kawamoto K, Del Fiol G, Strasberg HR, *et al.* Multi-national, multi-institutional analysis of clinical decision support data needs to inform development of the HL7 virtual medical record standard. *AMIA Annu Symp Proc.* 2010;2010:377–81.

28. Gray KA, Daugherty LC, Gordon SM, *et al.* Genenames.org: The HGNC resources in 2013. *Nucleic Acids Res.* 2013;41:D545–52. doi:10.1093/nar/gks1066

29. dbSNP Home Page. http://www.ncbi.nlm.nih.gov/SNP/. Accessed 3 Feb 2014.

30. Horaitis O, Cotton RGH. The challenge of documenting mutation across the genome: The human genome variation society approach. *Hum Mutat.* 2004;23:447–52. doi:10.1002/humu.20038

31. Logical Observation Identifiers Names and Codes (LOINC®) — LOINC. http://loinc.org/. Accessed 3 Feb 2014.

32. Lynch syndrome. http://ghr.nlm.nih.gov/condition/lynch-syndrome. Accessed 10 Mar 2014.

33. Lynch Syndrome Management. http://genefacts.org/index.php?option=com_content&view=article&id=492:management&catid=113:lynch-syndrome&Itemid=665. Accessed 10 Mar 2014.

34. Marquez E, Geng Z, Pass S, *et al.* Implementation of routine screening for Lynch syndrome in university and safety-net health system settings: Successes and challenges. *Genet Med.* 2013;15:925–32. doi:10.1038/gim.2013.45

35. Levy S, Sutton G, Ng PC, *et al.* The diploid genome sequence of an individual human. *PLoS Biol*. 2007;5:e254. doi:10.1371/journal.pbio.0050254

36. Peterson T, Doughty E, Kann MG. Towards precision medicine: Advances in computational approaches for the analysis of human variants. *J Mol Biol.* 2013;425:4047–63. doi:10.1016/j.jmb.2013.08.008

37. Riggs ER, Wain KE, Riethmaier D, *et al.* Towards a universal clinical genomics database: The 2012 International Standards for Cytogenomic Arrays Consortium Meeting. *Hum Mutat*. 2013;34:915–9. doi:10.1002/humu.22306

38. New NIH-funded resource focuses on use of genomic variants in medical care. http://www.nih.gov/news/health/sep2013/nhgri-25.htm. Accessed 8 Oct 2013.

39. Human genome overview - Genome Reference Consortium. http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/. Accessed 13 Mar 2014.

40. Bates D, Kuperman G. Ten Commandments for effective clinical decision support: Making the practice of evidence-based medicine a reality. *J Am Med Informatics Assoc*. 2003;10:523–30. doi:10.1197/jamia.M1370.

41. Regulations.gov - Proposed Rule Document.

http://www.regulations.gov/#!documentDetailD=HHS-OS-2014-0002-0001. Accessed 12 Mar 2014.

1. Maintain separation of primary molecular observations from the clinical interpretations of those data
2. Support lossless data compression from primary molecular observations to clinically manageable subsets
3. Maintain linkage of molecular observations to the laboratory methods used to generate them
4. Support compact representation of clinically actionable subsets for optimal performance
5. Simultaneously support human-viewable formats and machine-readable formats in order to facilitate implementation of decision support rules
6. Anticipate fundamental changes in the understanding of human molecular variation
7. Support both individual clinical care and discovery science

Figure 5.1: Desiderata for the integration of genomic data into EHRs described by Masys *et al.*

8. CDS knowledge must have the potential to incorporate multiple genes and clinical information.
9. Keep CDS knowledge separate from variant classification.
10. CDS knowledge must have the capacity to support multiple EHR platforms with various data representations with minimal modification.
11. Support a large number of gene variants while simplifying the CDS knowledge to the extent possible.
12. Leverage current and developing CDS and genomics infrastructure and standards.
13. Support a CDS knowledge base deployed at and developed by multiple independent organizations.
14. Access and transmit only the genomic information necessary for CDS.

Figure 5.2: Additional desiderata for the technical integration of WGS with CDS described by Welch *et al.*

The patient is a 32 years old male, with no personal history of colon cancer and no prior colonoscopies. The patient does not know his family health history. This patient previously had his genome sequenced, and this WGS information is available for assessment by CDS. This patient has a pathogenic mutation in the MLH1 gene, a gene associated with a high risk for Lynch syndrome. The clinical recommendation established by his healthcare organization is to recommend that patients over 20 years old with pathogenic mutations in the MLH1 gene receive a recommendation if one has not been performed in the past one to two years.

Figure 5.3: High-risk disease risk assessment clinical scenario

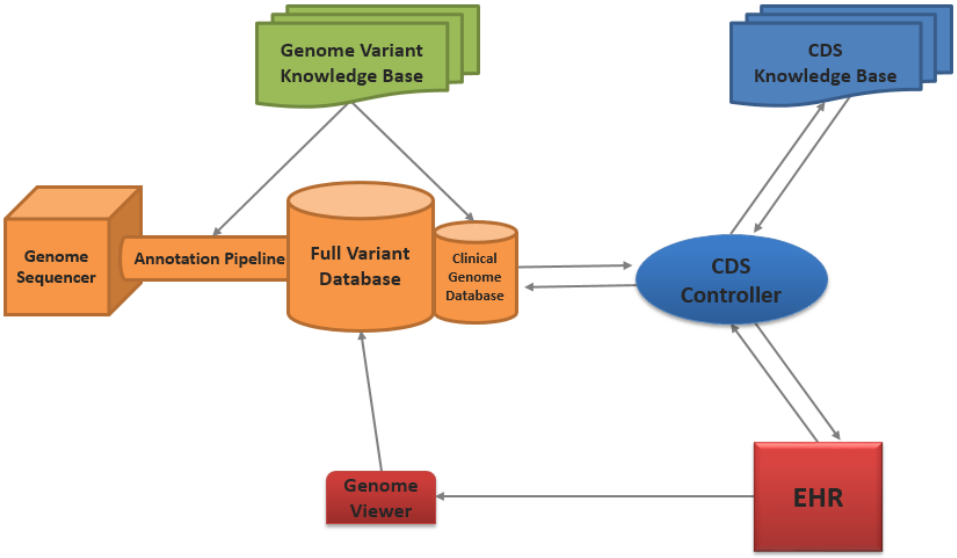


Figure 5.4: Overview of proposed CDS architecture for WGS information implemented as a prototype in this paper

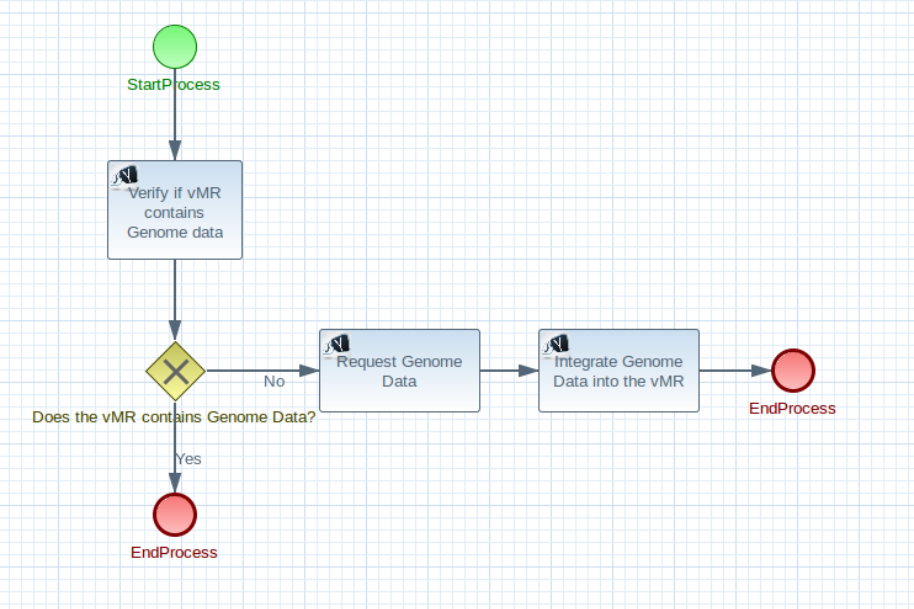


Figure 5.5: A visual diagram of the BPMN used within the VerifyGenome Data component of the SCA

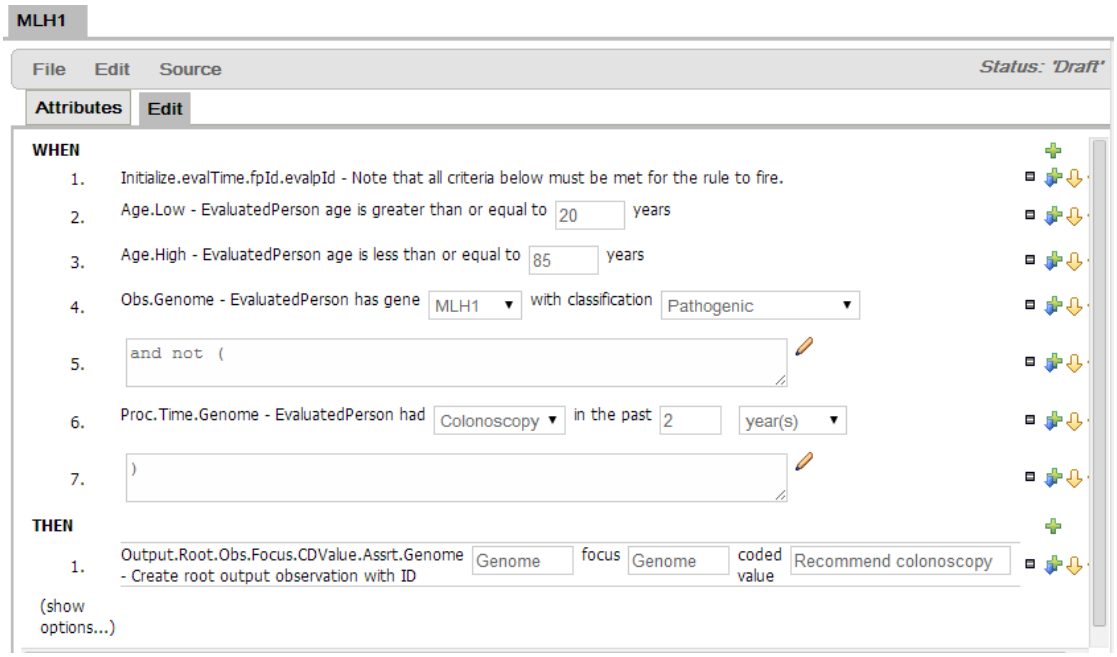


Figure 5.6: A screenshot of the MLH1 business rule developed in OpenCDS

```
<observationResult>
    <id root="1234" extension=""/>
    <observationFocus code="MLH1" codeSystem="genenames.org" codeSystemName='HGNC' displayName='MLH1'/>
    <observationValue>
        <text value="NM_000249.3:c.982C>T"
    </observationValue>
    <interpretation code="LA6668-3" codeSystem="2.16.840.1.113883.6.1" codeSystemName="LOINC" displayName="pathogenic"/>
</observationResult>
```

Figure 5.7: A vMR representation of a pathogenic *MLH1* genome variant

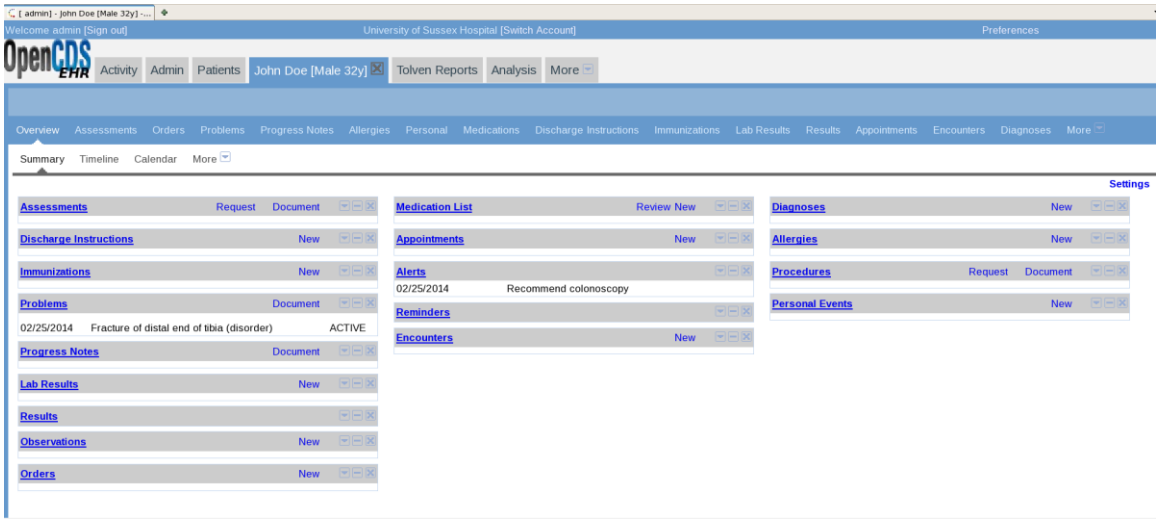


Figure 5.8: A screenshot of the Tolven EHR with the CDS recommendation generated by the architecture prototype

Table 5.1: Performance results of the architecture components using LoadUI

Component	Simultaneous Users	Total requests handled	Min request time (time in ms)	Max request time (time in ms)	Average request time (time in ms)	Standard deviation
Genome app	100	3109	25	697	**40**	47.77
OpenCDS	100	3015	7	914	**12**	17.04
Overall	20	650	356	4243	**944**	621.04

CONCLUSION

The previous chapters of this dissertation outline a body of work that represents an advance in the field of biomedical informatics in the domain of clinical decision support (CDS) for whole genome sequence (WGS) information. Chapter 1 represented a systematic review of the literature in the CDS for genetically-guided personalized medicine domain which identified a gap in published research in the area of CDS for WGS information. Chapter 2 described the need and justification for researching and developing CDS capabilities for WGS information. Chapter 3 built on previous work by Masys *et al.* entitled "Technical desiderata for the integration of genomic data into Electronic Health Records" by extending the desiderata with additional requirements specifically for CDS. Chapter 4 laid out the description and justifications for a proposed CDS architecture capable of supporting WGS information while adhering to the desiderata requirements. Finally, Chapter 5 described the efforts to build and test a functional prototype of this architecture.

Given the early stage of research on CDS for WGS, and the pending demand for such solutions in health care, this body of work has the potential to shape future research and development in this domain. Indeed, the work completed in this effort provides a framework and demonstration of feasibility which future research and development can build upon and improve. It is not the intention of this work to stand alone in its capabilities; rather it is anticipated that others will perfect and expand features described herein.

Certainly, it will require the cooperation and interdependence of several independent entities and organizations to realize the full potential of the approaches described in this work. To that end, we look forward to future research by ourselves and others to fully realize the potential of WGS-guided personalize medicine through the use of CDS capabilities.

APPENDIX A

MEMBERS OF THE CORE DESIDERATA PANEL

Brandon M. Welch, M.Sc. is a Biomedical Informatics Ph.D. candidate studying clinical decision support for genomics and family health history. Brandon received a Master's degree in Human Genetics from Tulane University. He was formerly a Vice President of InformedDNA, Inc., a nationwide genetic counseling provider, and has experience developing family health history and genetics applications.

Karen Eilbeck, M.Sc., Ph.D. is an associate Professor of Biomedical Informatics is focused on genomic annotation and knowledge representation for biology, and brings experience from the first human genome sequence, the Gene Ontology Consortium and the Sequence Ontology Project. She has developed file specifications for sharing of genomic annotations, including variant data and is currently working with the NIH driven ClinGen Initiative.

Guilherme Del Fiol, M.D., Ph.D. is an assistant Professor of Biomedical Informatics and has over 15 years of experience in health care informatics, particularly designing, implementing, and evaluating CDS interventions to improve the quality and safety of health care. He is a co-chair of the Clinical Decision Support Work Group at Health Level 7 (HL7). For 10 years he has been leading the development of international standards for CDS, particularly the Infobutton Standard, and open source software components to foster the wide dissemination of CDS interventions.

Laurence J. Meyer, M.D., Ph.D. is a molecular geneticist and practicing physician with board certifications in dermatology, internal medicine and genetics. As National Director

for Genomic Medicine at the Veterans Administration Medical Center, Dr. Meyer is working to deliver clinical genetics information to the primary and specialty care setting.

Kensaku Kawamoto, M.D., Ph.D. is the Associate Chief Medical Information Officer at the University of Utah Health Sciences Center and Assistant Professor of Biomedical Informatics. Dr. Kawamoto is the founder of OpenCDS, a multi-institutional open source CDS collaborative, and is a co-chair of the Clinical Decision Support Work Group at HL7. He also currently serves as the Initiative Coordinator of Health eDecisions, which is developing CDS guidelines for Meaningful Use requirements.

APPENDIX B

DESIDERATA SURVEY INSTRUMENT

Confidential

Validating a proposed technical desiderata for the integration of genomic information with clinical decision support (CDS) among domain experts

IRB INFORMED CONSENT COVER PAGE

BACKGROUND

The purpose of this research study is to obtain qualitative feedback on a set of requirements developed by a panel of experts for clinical decisions support (CDS) systems capable of supporting whole genome sequencing (WGS) information at the point of care. This is a research study being conducted by Brandon M Welch, M.S. of the University of Utah as part of his Ph.D. dissertation research. Your participation in this study is voluntary.

STUDY PROCEDURE

This survey will be anonymous and should take approximately 10-20 minutes to complete. You will be asked to identify your domain of expertise (genomics or CDS) and then review a set of seven requirements and provide free-form comments, suggestions, or other relevant feedback for each. You may choose to be identified for follow up questions regarding your comments from the panel of experts.

RISKS

No reasonably foreseeable risks are associated with this anonymous survey.

BENEFITS

We anticipate that the findings from this study could facilitate the enabling of CDS for WGS data and thereby eventually enable improved health and health care.

CONFIDENTIALITY

This survey will be anonymous by default. However, you may choose to provide your email to be re-contact regarding your responses. Any individual responses will be made anonymous in any publication. The research data will be stored on password protected computers.

PERSON TO CONTACT

If you have questions, complaints or concerns about this study, you can contact Brandon M Welch, MS at 585-455-0461. If you feel you have been harmed as a result of participation, please call Brandon M Welch, MS at 585-455-0461 who may be reached during regular business hours.

Institutional Review Board: Contact the Institutional Review Board (IRB) if you have questions regarding your rights as a research participant. Also, contact the IRB if you have questions, complaints or concerns which you do not feel you can discuss with the investigator. The University of Utah IRB may be reached by phone at (801) 581-3655 or by e-mail at irb@hsc.utah.edu.

Research Participant Advocate: You may also contact the Research Participant Advocate (RPA) by phone at (801) 581-3803 or by email at participant.advocate@hsc.utah.edu.

VOLUNTARY PARTICIPATION

It is up to you to decide whether to take part in this study. Refusal to participate or the decision to withdraw from this research will involve no penalty or loss of benefits to which you are otherwise entitled. This will not affect your relationship with the investigator.

COSTS AND COMPENSATION TO PARTICIPANTS

There are no costs or compensation associated with the study.

CONSENT:

By proceeding with the survey, I confirm that I have read this consent document and have had the opportunity to ask questions. I agree to participate in this research study and authorize the use and disclosure of the data I provide in this survey, as explained on this page.

1) To be eligible for this survey, you must be at least 18 years old and have domain expertise in genomics and/or clinical decisions support. Do you meet this eligibility criteria?

☐ Yes
☐ No
(If NO, please stop and exit the survey.)

BACKGROUND

The application of whole genome sequencing (WGS) in routine clinical care is quickly approaching. Having a patient's genomic information could allow clinician's to identify health risks, improve diagnosis, and tailor treatment, all under the paradigm of personalized medicine. However, were the WGS to be widely available in the clinic today, many of these capabilities could remain unachieved for a number of reasons, including the complexity of genetic analysis, practicing clinicians' limited proficiency in genetics, and the insufficient number of genetics professionals in the workforce (1).

Nevertheless, computerized clinical decision support (CDS) within the electronic health record (EHR) offers a reasonable solution to overcome these barriers and help clinicians leverage a patient's WGS information at the point of care (2). CDS entails providing clinicians, patients, and other healthcare stakeholders with pertinent knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and healthcare (3). Examples of point-of-care CDS include medication dosing support, order facilitators, alerts and reminders, relevant information display, expert systems, and workflow support (4).

Masys et al. developed a set of requirements to incorporate genomic variation into the EHR for the provision of healthcare services (5). The requirements include the following:

1. Maintain separation of primary molecular observations from the clinical interpretations of those data
2. Support lossless data compression from primary molecular observations to clinically manageable subsets
3. Maintain linkage of molecular observations to the laboratory methods used to generate them
4. Support compact representation of clinically actionable subsets for optimal performance
5. Simultaneously support human-viewable formats and machine-readable formats in order to facilitate implementation of decision support rules
6. Anticipate fundamental changes in the understanding of human molecular variation
7. Support both individual clinical care and discovery science

This set of requirements provides a strong set of guiding principles to integrate genomic information into the EHR. However, in our effort in developing CDS capabilities for WGS data, we have identified ADDITIONAL REQUIREMENTS specific to CDS to augment the original Masys desiderata. These additional requirements are as follows (further described in the survey):

8. Keep CDS knowledge separate from variant classification
9. Support a large number of gene variants while keeping CDS logic as simple as possible
10. CDS knowledge must be able to incorporate multiple genes and clinical information
11. Access and transmit only the genomic information necessary for CDS
12. Leverage current and developing CDS and genomics infrastructure and standards
13. Support a CDS knowledge base deployed at and developed by multiple independent organizations
14. Have the capacity to support multiple EHR platforms with various data representations with minimal modification

OBJECTIVE

The intent of this survey is to VALIDATE THIS ADDITIONAL SET OF REQUIREMENTS AMONG DOMAIN EXPERTS in CDS and/or genomics. On subsequent pages of this survey, each of these requirements are described in further detail with a Likert scale question to assess the importance of that particular requirement. If desired, there is an optional comment field after each requirement to provide suggestions for improvement and/or reasoning for your response on the importance of the requirement. Finally, general comments can be provided regarding any overarching recommendations or ideas for additional requirements. Survey responses will be reviewed and incorporated into the requirements by the following panel of experts:

Brandon M Welch, MS
Karen Eilbeck, MS, PhD
Laurence J. Meyer, MD, PhD
Guilherme Del Fiol, MD, PhD
Kensaku Kawamoto, MD, PhD

We request that you keep the content confidential until survey results have been published. Thank you for your participation.

DEFINITIONS OF SELECT TERMS USED IN THE SURVEY

Molecular observations: The patient-specific information obtained from sequencing technologies which include the identification and location of variants in DNA, RNA, and protein sequence. Example representation in Human Genome Variation Society (HGVS) format: MSH2 c.1452_1455delAATG

Variant interpretations: Disease causing genes can have one of many known variants present. However, not all variants are pathogenic. Some variants are benign, and others have unknown pathogenicity (also known as variants of unknown significance [VUS]). The American College of Medical Genetics and other groups have put forth various recommendations for variant classification, including pathogenic, likely pathogenic, likely benign, benign, or variant of unknown significance (18). An example variant classification for the MSH2 variant in the above example is Pathogenic.

CDS knowledge: CDS (described above) provides guidance to clinician on how to care for a patient when a particular variant is present or absent. Continuing the previous example of a patient with a pathogenic mutation in the MSH2 gene, a CDS alert for this person could say the following: "This patient is at increased risk for colorectal cancer (Lynch syndrome I). The patient should be placed on an NSAID and receive a colonoscopy annually. This patient should be referred for genetic counseling and to an oncology specialist."

REFERENCES

1- http://www.ncbi.nlm.nih.gov/pubmed/22918138
2- http://www.ncbi.nlm.nih.gov/pubmed/22922173
3- http://www.ncbi.nlm.nih.gov/pubmed/17213487
4- http://www.ncbi.nlm.nih.gov/pubmed/21415065
5- http://www.ncbi.nlm.nih.gov/pubmed/22223081

2) What type of EXPERTISE do you have? Select the field with which you are most closely associated. (select all that apply)

- ☐ Genomics
- ☐ Clinical Decision Support

ADDITIONAL REQUIREMENT #1
Keep CDS knowledge separate from the variant classification.

EXPLANATION
Masys et al. describe the importance of separating molecular observations (e.g. DNA sequence) from clinical interpretation (e.g. variant classifications) due to the need to update variant interpretation as knowledge changes and grows over time. To illustrate, one study found that over a seven year period, 14.5% of reported variant classifications had to be reclassified (1). Likewise, it is essential to separate CDS knowledge from both molecular observations and variant interpretations. CDS must have the ability to manage evolving and frequently changing gene variant interpretations efficiently without requiring changes to the underlying CDS knowledge each time a variant's classification changes. Separation of CDS knowledge from variant interpretation allows CDS knowledge to be more efficiently handled and maintained.

1-http://www.ncbi.nlm.nih.gov/pubmed/22481129

3) In your opinion, how important is this requirement?

- ☐ Very important
- ☐ Important
- ☐ Neither important nor unimportant
- ☐ Unimportant
- ☐ Very unimportant
- ☐ Unsure/No opinion

4) Comments regarding this requirement:

(If desired, please provide suggestions for improvement f

ADDITIONAL REQUIREMENT #2

Support a large number of gene variants while keeping CDS logic as simple as possible.

EXPLANATION

There are roughly 1200 known variants in the adenomatous polyposis coli (APC) gene, a gene linked to a rare form of colon cancer (1). Likewise, there are nearly 2000 known variants in the cystic fibrosis gene CFTR (2). Given the potentially high number of variants per gene, it may be inefficient to create CDS logic for every known variant in each disease-causing gene. Furthermore, as novel variants are discovered, it will be difficult to update CDS logic for every gene variant that is discovered. Therefore, to manage this complexity, variants with the same or similar clinical impact should be classified accordingly. CDS logic can then be simplified by developing logic which leverages the variant interpretation rather than the specific variant. Nevertheless, in cases where a particular variant has a unique and clinically important impact or where machine learning CDS models could utilize individual variants, genetic information at the variant level should still be accessible to CDS logic. In summary, CDS logic can be greatly simplified by classifying variants into groups of common clinical impact, while still supporting inferencing at the individual variant level where necessary.

1-http://chromium.liacs.nl/LOVD2/colon_cancer/home.php?select_db=APC
2-http://www.genet.sickkids.on.ca/StatisticsPage.html

5) In your opinion, how important is this requirement?

- ☐ Very important
- ☐ Important
- ☐ Neither important nor unimportant
- ☐ Unimportant
- ☐ Very unimportant
- ☐ Unsure/No opinion

6) Comments regarding this requirement:

(If desired, please provide suggestions for improvement f

ADDITIONAL REQUIREMENT #3
CDS knowledge must be able to incorporate multiple genes and clinical information.

EXPLANATION
A relatively small number of Mendelian diseases, such as cystic fibrosis and sickle-cell anemia, are affected by variants within a single gene responsible for producing the characteristic phenotype. As a result, such cases are fairly straightforward to assess. With nearly every human condition affected one way or another by a genetic influence, most diseases, in particular common diseases, are caused or affected by multiple genetic influences and environmental factors. For example, there are potentially hundreds of genetic loci contributing to type 2 diabetes risk (1). In order to provide an accurate risk assessment and decision support, all relevant genetic loci need to be considered, as well as any relevant clinical factors (e.g. age, weight, health history, co-morbidities) and environmental influences (e.g. diet, physical activity, stress). Therefore, CDS for the WGS must have the capacity to leverage and incorporate information from multiple genomic and non-genomic data sources.

1- http://www.ncbi.nlm.nih.gov/pubmed/21778616

7) In your opinion, how important is this requirement?

- ☐ Very important
- ☐ Important
- ☐ Neither important nor unimportant
- ☐ Unimportant
- ☐ Very unimportant
- ☐ Unsure/No opinion

8) Comments regarding this requirement:

(If desired, please provide suggestions for improvement f

ADDITIONAL REQUIREMENT #4
Access and transmit only the genomic information necessary for CDS.

EXPLANATION
The separation of CDS knowledge from molecular observations and variant interpretations will require relevant genetic information being accessed and sent to a CDS rules engine (or equivalent) for processing. It will be inefficient and insecure to transmit an entire genome file for every CDS rule. The processing capacity required to transmit and sift through an entire genome for CDS knowledge will hinder the ability to provide CDS at the point of care in real-time. Furthermore, HIPAA requires that only the minimum protected health information needed to satisfy a particular purpose or carry out a function be used or transmitted (1). Therefore, a CDS architecture must only transmit the relevant genes and any associated molecular observations and variant interpretations.

1-
http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/minimumnecessary.html

9) In your opinion, how important is this requirement?

- ☐ Very important
- ☐ Important
- ☐ Neither important nor unimportant
- ☐ Unimportant
- ☐ Very unimportant
- ☐ Unsure/No opinion

10) Comments regarding this requirement:

(If desired, please provide suggestions for improvement f

ADDITIONAL REQUIREMENT #5
Leverage current and developing CDS and genomics infrastructure and standards.

EXPLANATION
Both the CDS and genomics fields have benefited from extensive research and development over the years. Indeed, both fields have well developed infrastructure and standards to support its uses. Therefore, it is important to leverage these standards and infrastructure. Examples include using HGVS and/or dbSNP to represent specific molecular observations; ACMG recommendations for variant classification; HL7 Clinical Genomics standards for the representation of genetic information; Arden Syntax, GELLO, GLIF3, GEM, or the HL7 CDS Knowledge Artifact Implementation Guide for CDS knowledge representation; the HL7 Decision Support Service standard for delivering CDS as a service; and open-source, standards-based resources such as OpenCDS. While many of these standards and resources may not be completely sufficient for meeting the needs of CDS for WGS, it still represents significant relevant effort. It will be important to leverage current and developing CDS and genomics infrastructure and standards.

11) In your opinion, how important is this requirement?

- ☐ Very important
- ☐ Important
- ☐ Neither important nor unimportant
- ☐ Unimportant
- ☐ Very unimportant
- ☐ Unsure/No opinion

12) Comments regarding this requirement:

(If desired, please provide suggestions for improvement f

ADDITIONAL REQUIREMENT #6

Support a CDS knowledge base deployed at and developed by multiple independent organizations.

EXPLANATION

With the potential for genomic information to impact nearly every clinical decision and the clinical application of genomics rapidly evolving, the time and cost for a single entity or organization to manually create and update CDS knowledge will be prohibitive. Indeed, no one organization will be able to author and manage all CDS knowledge for all WGS use cases. Furthermore, there must be an efficient and scalable mechanism to consistently modify CDS knowledge everywhere it is deployed. Ideally, a standardized CDS infrastructure would allow multiple organizations, entities, or individuals to create, publish, and distribute CDS knowledge efficiently to multiple consuming health care organizations. Such an approach will allow a specialized entity (e.g. pharmacogenomics experts) to develop and manage CDS knowledge and subsequently distribute to 'subscribing' health care organizations. With an ecosystem of independently developed CDS knowledge available, it becomes more feasible for health care organizations to have the most up to date and accurate CDS knowledge for the entire genome in an affordable way.

13) In your opinion, how important is this requirement?

- ☐ Very important
- ☐ Important
- ☐ Neither important nor unimportant
- ☐ Unimportant
- ☐ Very unimportant
- ☐ Unsure/No opinion

14) Comments regarding this requirement: (If desired, please provide suggestions for improvement f

ADDITIONAL REQUIREMENT #7

Have the capacity to support multiple EHRs platforms with various data representations with minimal modification.

EXPLANATION

The reality of the health information environment in the US today is that multiple healthcare organizations use multiple EHR and health information management systems (1). Often, these health information management solutions store and represent the same health information differently. This can be a challenge when trying to harness the information within different health IT systems in different organizations to provide CDS. With the need to distribute and share WGS enabled CDS knowledge across multiple organizations (see previous requirement), the CDS architecture would ideally be EHR agnostic, where a CDS rule can be written once, then run consistently anywhere. A number of initiatives aimed at supporting this effort are underway, including the Health eDecisions initiative, OpenCDS, the SMART platform, and the CDS Consortium, to name a few.

1-
http://www.softwareadvice.com/medical/electronic-medical-record-software-comparison/#buyers-guide

15) In your opinion, how important is this requirement?

- ☐ Very important
- ☐ Important
- ☐ Neither important nor unimportant
- ☐ Unimportant
- ☐ Very unimportant
- ☐ Unsure/No opinion

16) Comments regarding this requirement: (If desired, please provide suggestions for improvement f

Thank you for participating in this survey.

17) If appropriate, please provide any ADDITIONAL GENERAL COMMENTS, including additional concepts or requirements not described in this survey that you feel is important.

_____

18) Your survey responses are anonymous. However, if you would be willing to be RE-CONTACTED regarding one or more of your responses or to provide additional feedback to study coordinators, please provide your email address.

_____

APPENDIX C

GLOSSARY OF TERMS USED

CDS controller: A component of a SOA architecture which links several services together.

CDS knowledge: A representation of clinical knowledge in the form of logic, rules, expressions, guidelines or algorithms.

CDS knowledge base: A repository of CDS knowledge.

Clinical genome database: A repository which stores only variants of known or potential clinical importance.

Clinical interpretation: The clinical impact of a variant.

Full genome database: A repository which stores all variants of an individual's genome.

Genome annotation: The process of locating and identifying key features of a genome.

Genome interpreter: A visual interface which allows a clinician to manually review a patient's genome variants.

Genome sequencing: The process of obtaining the DNA sequence of an individual.

Genome variant knowledge base: A repository of variants and associated clinical interpretation.

Genome variant: A difference in a genome relative to a reference genome sequence.

Service: A self-contained component with well-defined, understood capabilities.

Service oriented architecture (SOA): A software design methodology which contains several independent services

APPENDIX D

COMPUTER CODE AND CONFIGURATION USED TO DEVELOP THE PROTOTYPE ARCHITECTURE

SQL code to recreate database tables

```
CREATE DATABASE 'genomedb'

CREATE TABLE `patient_genome` (
`MRN` INT(10) NULL DEFAULT NULL,
`gene` VARCHAR(50) NULL DEFAULT NULL,
`refseq` VARCHAR(50) NULL DEFAULT NULL,
`nuc_var` VARCHAR(100) NULL DEFAULT NULL,
`pro_var` VARCHAR(100) NULL DEFAULT NULL,
`interpretation` VARCHAR(100) NULL DEFAULT NULL,
`id` INT(10) NOT NULL AUTO_INCREMENT,
PRIMARY KEY (`id`)
)
COMMENT='stores patient-specific genetic information will be queried by
the web service'
ENGINE=InnoDB;

CREATE TABLE `gene_variant_table` (
`ClinVarSetID` INT(11) NOT NULL,
`interpretation` VARCHAR(45) NULL DEFAULT NULL,
`nucleotide_variant` VARCHAR(90) NULL DEFAULT NULL,
`gene_name` VARCHAR(30) NULL DEFAULT NULL,
PRIMARY KEY (`ClinVarSetID`)
)
COMMENT='stores gene variants from ClinVar, used to update
interpretations in patient_genome table'
```

Download ClinVar XML data

$ wget ftp://ftp.ncbi.nlm.nih.gov/pub/clinvar/xml/ClinVarFullRelease_2014-01.xml.gz

$ gunzip ClinVarFullRelease_2014-01.xml.gz

XSLT code to transform ClinVar XML to database import XML

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<xsl:stylesheet version="1.0"
xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:template match="/">
      <list>
      <xsl:text>&#xa;</xsl:text>
      <xsl:for-each select="ReleaseSet/ClinVarSet">
            <xsl:for-each
select="ReferenceClinVarAssertion/MeasureSet/Measure/AttributeSet/Attrib
ute[@Type='HGVS, coding, RefSeq' or @Type='HGVS, protein'] |
eferenceClinVarAssertion/MeasureSet/Measure/AttributeSet/XRef[@DB='dbSNP
']">
            <row>
                  <field name='ClinVarSetID'><xsl:value-of
select="../../../../../@ID"/></field>
                  <field name='interpretation'><xsl:value-of
select="../../../../ClinicalSignificance/Description"/> </field>
                  <xsl:if test="@Type='HGVS, coding, RefSeq' or
@Type='HGVS, protein'">
                        <field name='{@Type}'><xsl:value-of
select="."/></field>
                  </xsl:if>
                  <xsl:if test="@DB='dbSNP'">
                        <field name='dbSNP'>rs<xsl:value-of
select="@ID"/></field>
                  </xsl:if>
                  <field name='gene_name'><xsl:value-of
select="../../../Measure/MeasureRelationship/Symbol/ElementValue"/></fie
ld>
            </row>
            <xsl:text>&#xa;</xsl:text>
                  </xsl:for-each>
            </xsl:for-each>
            </list>
      </xsl:template>
</xsl:stylesheet>
```

Program used to run XSLT on ClinVar XML file

```
$ xsltproc -o [output_file] [stylesheet_file] [input_file]
```

SQL import CSV to database code

```
LOAD XML LOCAL INFILE
'C:\\Users\\bmw\\SkyDrive\\Documents\\PhD\\Thesis\\Project\\ClinVar\\Cli
nVarOutput-final.xml' INTO TABLE genomedb.genomekb;
```

Update Variant Interpretation from GenomeKB Table

```
UPDATE IGNORE patient_genome
SET interpretation =(SELECT genomekb.interpretation FROM genomekb
WHERE patient_genome.gene = genomekb.gene
AND (patient_genome.refSNP = genomekb.refSNP
OR  patient_genome.nuc_var = genomekb.nuc_var;
```

Location of the CDS knowledge developed in OpenCDS

http://tolven.chpc.utah.edu:8081/drools-guvnor/org.drools.guvnor.Guvnor/Guvnor.jsp

LOINC terms added to the Apelon DTS instance

53037-8	Genetic disease sequence variation interpretation	1	Pathogenic	LA6668-3
		2	Presumed pathogenic	LA6669-1
		3	Unknown significance	LA6682-4
		4	Benign	LA6675-8
		5	Presumed benign	LA6674-1

Code used to create the MLH1 mutation in the test patient's genome

```
insert into patient_genome
values ('33333', 'MLH1', null, 'NM_000249.3:c.982C>T',
'NP_000240.1:p.Gln328Ter', null, default);
```